THE SELECTION OF WORKING DOGS

'The use of impulsivity and core affect to optimize the potential selection of working dogs'

A thesis submitted in partial fulfilment of the requirements of the University of Lincoln for the degree of Doctor of Philosophy

Karen Linda Brady

Supervisors: Daniel Mills, Helen Zulch, Nina Cracknell, Hannah Wright

APRIL 1, 2017
UNIVERSITY OF LINCOLN
Karen Brady

Abstract

Police and military working dogs are highly trained to be able to perform a variety of roles but withdrawal from service due to behavioural problems is an issue (Evans *et al*, 2007). To achieve optimal field performance there is a need to be able to assess resilience in dogs and understand factors that impact performance, such as arousal and distractibility. Relationships between arousal and performance were examined using a scent detection task in a population of pet dogs and highlighted the challenges in assessing arousal in terms of physiological measures (heart rate variability). This line of enquiry was not pursued further and the focus was shifted to developing methods to assess temperament traits which may be linked to withdrawal from service. A review of published tests that assessed behavioural characteristics relevant to working dogs revealed a lack of reliable methods for assessing behavioural characteristics relevant to working dogs, and unclear predictive ability in terms of service performance outcomes following selection and certification. Working dogs need to be resilient enough to cope with working environments that they are likely to encounter, but also they need to be able to work despite distractions. Impulsivity is relevant to the assessment of distractibility and while questionnaire measures (DIAS, Wright *et al,* 2011) can be used to assess the trait in dogs, existing behavioural measurement methods for impulsivity require extensive training (Wright *et al,* 2012). This lead to the development of a simplified behavioural test to assess impulsivity using a spatial discounting paradigm in dogs over the age of 2 years which found that more impulsive individuals travel a shorter distance for a larger reward before switching to a small reward, and less impulsive individuals travelled further. Resilience and distractibility also depend partly on sensitivity to rewards (positive distractions - temptations) and aversives (negative distractions – anxieties) which can be assessed psychometrically in dogs from 10 weeks of age using the Positive and Negative Activation Scale - PANAS, Shepherd and Mills 2002). The PANAS and the DIAS were used to collect data within the UK police and military dog sector and found that police dogs that had been withdrawn from service for behavioural reasons were found to score significantly lower for “Responsiveness” in terms of impulsivity (using the DIAS), and Positive Activation “Energy & Interest” in terms of core affect (using the PANAS) compared to police dogs in active service. In a working dog setting questionnaires cannot always be relied upon so behavioural tests were developed to assess these elements. Dogs with temperament profiles similar to the active working dogs acquired a new task in fewer trials compared to the withdrawn group but task acquisition was statistically similar at retest. These results suggested that it may be the initial learning processes that are important in test performance and may be indicative of individuals at risk of withdrawal from field service. Further information is still needed in terms of military working dog withdrawal and also to establish if the temperament profiles observed in withdrawn dogs are a result of them failing in their work, or if the profile is responsible for their poor performance leading to withdrawal.

Contents

Chapter 1: Introduction

1.1 Individual Differences:

Being an individual is how we as humans define ourselves. There are a number of factors explaining potential reasons for our individuality from our genetics to our belief systems (Maltby *et al*, 2013). Who we are as an individual is thought to contribute to our life choices as well as our goals and ambitions, and whether or not we achieve these (Maltby *et al*, 2013). Within the human literature individual differences are generally referred to as personality, but this is a term that is sometimes avoided within the animal literature, reasons given often include a desire to avoid anthropomorphism (Gosling, 2008). Even within the human psychology literature differences exist within definitions of personality, mainly over the impact of genetics or the environment on shaping it, and the flexibility of personality traits (Gosling, 2008). In general terms personality has been defined as 'characteristics of individuals that describe and account for temporally stable patterns of affect, cognition, and behaviour' (Gosling, 2008). Some researchers within the field of animal science have taken on this term (Fidler *et al*, 2007. Wolf *et al*, 2007), although consensus over what it means tends to vary between disciplines and often terms such as behavioural syndromes (favoured within ecology and used to refer to collections of behaviours observed within or between contexts which is arguably similar to personality definitions (Sih *et al*, 2004)) and temperament (often used within animal behaviour although with an emphasis on a biological basis for behaviour i.e. genetic inheritance (McCrae *et al*, 2000)) are used to refer to the same concept, which could be classed as personality or individual differences (Goldberg, 1993). In humans temperament is sometimes described as the biological basis of individual differences in behaviour and having an inherited component (McCrae *et al*, 2000), with personality building on this as individuality is developed through life experiences (Goldberg, 1993. Rothbart *et al*, 2000).

There is growing interest in and evidence to support the existence of individual differences in nonhuman animals (Gosling *et al*, 2002), and this has been studied in a number of species (Stamps, 2012) including fish (Huntingford, 1976), rodents (Koolhaas *et al*, 2001), birds (Verbeek *et al*, 1996), sheep (Romeyer & Bouissou, 1992), goats (Lyons *et al*, 1988), donkeys (French, 1993), cats (McCune, 2005), horses (Seaman *et al*, 2002), primates (Capitainio *et al*, 2005) and dogs (Taylor & Mills, 2006). These individual differences have been studied from a number of perspectives from ecology to behaviour (Momozawa *et al*, 2003. McDougall *et al*, 2005. Reale *et al*, 2007. Martin & Reale, 2008), with a view to understanding why individual differences exist and how they impact on an individual's interactions with the environment and with others.

1.2 Temperament, Behaviour, Character and Personality

There is a lack of consensus with regards to the best approach for assessing individual differences (Seaman *et al*, 2002). Unclear distinctions and the interchangeable use of terminology such as temperament, behaviour, and personality is a problem in both research and its applications (Deidrich & Giffroy, 2006).

In the study of individual differences in animal behaviour the terms temperament, character and personality are frequently used interchangeably. However, it may be useful to recognise subtle differences between these constructs (Horwitz & Mills, 2009). Throughout this thesis some of the key terms relevant to the study of individual differences will be used in the following ways unless directly quoting authors who use the term in another way;

- Temperament is the individual differences in behavioural responses, consistent across time and context that is grounded in emotional states, core affect and associated behavioural regulatory processes. It is evident from an early age (Deidrich & Giffroy, 2006). It may be described as the fundamental reason for observed differences in behaviour between individuals of the same species (Seaman et al, 2002) and should be an accurate predictor of behaviour in a range of contexts.
- Character is an individual's style of responding that is the product of the environment (Horwitz & Mills, 2009). It is the habitual patterns of behaviour formed as a result of experiences within the environment (Horwitz & Mills, 2009). In this context, character can be more greatly influenced by changing the environment (Horwitz & Mills, 2009) whereas temperament is founded more on the relatively "hard wired" outcomes of early life experiences and genetics, and so can be viewed as encompassing the way an individual "naturally" responds to its' environment (Deidrich & Giffroy, 2006).
- Personality is the sum of both temperament and character and manifests itself in the general responses to the environment, both learned and innate (Horwitz & Mills, 2009).

There are a number of tools used for assessing personality in humans, although many are built on a common theoretical Five Factor Model. The Five Factor Model is based on trait concepts usually referred to as extraversion (outgoing, confident), agreeableness (willingness), conscientiousness (being thorough and thoughtful), neuroticism (sensitive, obsessive, anxious) and intellect (ability to reason and understand, mental acuity) (Goldberg, 1993. Judge & Ilies, 2002. Chamorro-Premuzik & Furnham, 2003. Komarraju *et al*, 2009). Tests can be focused on all of these traits, a subsection of them, or based on components of the traits themselves, depending on the aspect of personality that is of interest (Derryberry & Reed, 1994. Rothbart *et al*, 1994). The study of animal personality, as with humans, is generally based around the notion of trait constructs and the Five Factor Model discussed has been evaluated for use in assigning trait constructs to a number of species (Gosling & John, 1999). The five factors of the human model have been found to exist albeit to a lesser extent in many species (Gosling & John, 1999), but assessment is challenging as the terms do not always apply across species due to their reliance on self-report. This has led to the development of alternative trait constructs in animals that reflect what might be considered more representative of temperament than personality as defined within this thesis. The study of animal behavioural traits is generally based around the notion of trait constructs and often involves measuring behaviour in a range of contexts and then inferring traits that explain the behavioural tendencies observed.

Temperament is adaptive in an ultimate biological sense. It comprises fundamental elements or traits and different combinations of traits may be more advantageous in different environments. Traits therefore naturally exist along a spectrum within a species since no single expression of a trait is consistently the best in all situations (Sheppard & Mills, 2003). Temperament has been studied to understand the impact it has on fitness in wild populations and the role natural selection has on trait development and the persistence of individual differences in natural populations (Stamps & Groothuis, 2010).

1.3 Behavioural Traits and performance

Individual behavioural differences (or "behavioural syndromes" as they are sometimes called in the behavioural ecology literature (Bell, 2007)) have been investigated in a number of species (Reale *et al*, 2007) and are linked to species level success in terms of survival and fitness in different ecological niches (Dingemase & Reale, 2005), but they are also of relevance at a more proximate level.

Behavioural traits are assigned to individuals through assessing behavioural responses over a number of contexts (Maejima *et al*, 2007). A behavioural response is the way an individual acts following exposure to a stimuli (event or object that elicits a reaction). While behavioural responses are of interest and can be useful in understanding the individual, by assessing behavioural responses to a number of stimuli, this can give a better indication of how that individual tends to respond to certain groups of stimuli (such as potentially threatening, or excitatory stimuli (Slabbert & Odendaal, 1999)). From this, behavioural traits such as confidence and anxiety can be measured, which are likely important in terms of performance (Palisin, 1986. Slabbert & Odendaal, 1999. Chamorro-Premuzik & Furnham, 2003. Coleman *et al*, 2005).

A number of traits have been studied in order to try to identify individual factors associated with success. Traits have been used to predict academic success in adults (Chamorro-Premuzik & Furnham, 2003) as well as in children (Palisin, 1986). In animals individual differences have been linked to learning factors such as training success in primates and understanding this relationship has been used to select animals suitable for training, research projects, and to identify individuals in need of alternative training methods or those unsuitable for work (Coleman *et al*, 2005). Furthermore, behavioural tests examining trait “confidence” in police dogs have been correlated with the success of the working dog in terms of certification (passing their working dog test (Slabbert & Odendaal, 1999)) and in detection dogs in Japan the temperament trait “desire for work” has been shown to correlate with certification outcome (Maejima *et al*, 2007).

Individual differences in behaviour have been used to understand social behaviour in human adults (Mann, 1959. Sanson *et al*, 2004), and children (Rothbart *et al*, 1994), particularly in terms of what makes a successful match of individuals to form cohesive pairs or social units, such as foster placements for human children (Doelling & Johnson, 1990), and pairing of laboratory primates (Capitanio *et al*, 2015). Individual differences have been found to be related to behavioural problems in human children (Sanson *et al*, 2004) and personality disorders in human adults (Fiedler *et al*, 2004). It can predict behaviour problems and potentially could be used to predict behavioural adjustment throughout the early years and the ability of humans to cope with (Palisin, 1986) and adapt to (Fiedler *et al*, 2004) change. Temperament in animals has been linked to coping ability with higher levels of trait confidence linked to more active coping styles and individuals possessing certain traits showing greater ability to adapt to environmental challenges (Seaman *et al*, 2002). The trait impulsivity has been linked to behaviour problems relating to a lack of self-control and frustration tolerance in dogs (Wright *et al,* 2011) and it is likely that the intensity of a range of temperament traits can be related to problem behaviour, which likely has a negative impact on success and/or performance.

At a practical level, being able to measure traits associated with successful or unsuccessful task performance could be used to enable the prediction of outcomes in a variety of contexts. Quick, simple and effective temperament tests would be an asset to those who work with animals, whether the tests are used in order to match animals to potential owners (e.g. van der Borg *et al*, 1991. De Palma *et al*, 1995), create successful social groups (e.g. Capitainio *et al*, 2005), or to select working dogs including disability assistance dogs, detection and protection dogs (e.g. Murphy, 1995. Knol *et al*, 1998).

1.4 Performance and Arousal

Performance may also be impacted by arousal. Individuals may have different levels of arousal at which they function at their best0 (Hagvet and Hanin, 2007). The sports science literature, in particular, has examined the impact arousal has on performance (Salminen *et al*, 1995. Hagvet and Hanin, 2007). There are two ways in which arousal may relate to performance. The Yerkes-Dodson Law hypothesises that different tasks have different optimal levels of arousal for peak performance

(Cohen, 2011), while the Individual Zone of Optimal Performance (IZOF) theory suggests that different people have different but specific optimal levels of arousal for peak performance (Hagvet and Hanin, 2007). The concept of Individual Zones of Optimal Functioning or "IZOF hypothesis" suggests that an athlete performs at their best while experiencing an ideal emotional state, often linked to level of anxiety (Hagvet and Hanin, 2007), and methods to manipulate arousal to enhance performance have been investigated with some success (Robazza *et al*, 2004). This could mean that arousal experienced prior to or even during an exercise may impact differently on the performance of individuals, this in turn could impact on our ability to predict outcomes; i.e. that there is no optimal level of arousal for the task, rather it is individually specific. While the Yerkes Dodson law has been examined in dogs performing a detour task and found that working dogs bred for low arousal completed the task faster than pet dogs classed as the high arousal group, the concept of Individual Zones of Optimal Functioning has not been examined in detail in working dogs (Bray *et al*, 2015). These theories are not necessarily mutually exclusive. For example it may be that in general low levels of arousal are best for cognitive tasks, but within that low level of arousal some individuals will be at the higher end of the scale to others (Hagvet and Hanin, 2007).

1.5 Working Dogs

A working dog is a dog trained to perform a specific task or tasks. This ranges from gun dogs trained to flush and retrieve on shoots, but more specifically within this thesis dogs trained for detection or protection work associated with the police and military. Protection dogs are trained to track suspects, property, or victims and to halt suspects and take them down if required. Detection dogs are trained to search for specific scents ranging from firearms, to narcotics, cash and explosives. Dogs have been shown to be better than machines in terms of scent detection (Furton & Myers, 2001) and they are frequently used within the military for both detection and protection roles, providing valuable support for military operations. Military working dogs have been described as vital assets in the war against terror (Tiffoli & Rolfe, 2006). Their olfactory capabilities make dogs well suited to detection work in a variety of settings from cadavers (Komar, 1999), avalanche victims (Tschirky et al, 2000), disaster survivors, lost persons, terrorists (Lorenzo et al, 2003), drugs (Lorenzo et al, 2003) and explosives (Gazit et al, 2003).

Potential military working dogs have a high initial cost and require extensive training, these dogs need to provide a high quality service for the longest time possible (Moore *et al*, 2001). The ability to perform their job effectively relies not only on initial training, but also on physical and mental preparedness for field conditions. Training aims to not only ensure the dog has sufficient ability to work, but also that the dog will be able to cope in a working environment which can be extremely challenging. For example, from a physical perspective dogs may be required to scale obstacles or treacherous terrain, while at a psychological level dogs may be required to work in very crowded, noisy and highly distracting as well as potentially threatening situations.

1.5.1 Selection Testing

Current selection tests to identify individuals suited to work in such environments vary within and between countries. Tests are based on research and modified for field use. Selection tests are varied across the working dog sector, tests are based around the scientific literature and practitioners often take elements from a number of recognised tests as part of their assessment process. Working dog selection tests can be carried out at 8 weeks of age, or at the time the dog is selected for work if centres are not breeding their own dogs. Selection aims to identify individuals with the ability to work as well as those that possess traits thought to be desirable in a working dog setting. Testing can vary, but in general the following applies:

Selection testing for police and military working dogs is usually conducted in a novel environment to assess how the dog responds to new places (Svartberg & Forkman, 2002. Svartberg, 2005), this is often used as part of the assessment for confidence, boldness, shyness and or fearfulness (Gosling *et al*, 2009) depending on the assessor (varied practice between centres in measurements taken). Various measures can be used such as the way the dog enters the room (as outlined in Svobodova *et al*, 2008), and the level of exploration based on the ground covered and objects approached (taken from Wilsson & Sundgren, 1998). While in the literature such tests are highly standardised, in practice environments are not always standardised or cleaned between each dog or cohort of dogs being tested due to practicality.

The tests for selection will include some form of sound sensitivity test, such as exposing the dog to a loud bang or gunshot and assessing their response and recovery. Metallic object noise sensitivity tests have been taken and modified from the literature (Svartberg, 2002, Svartberg, 2005, Svartberg *et al*, 2005), as well as gunshot tests which are not always practical to perform in the field (Ruefenacht *et al* 2002. Svartberg *et al*, 2005. Gosling *et al*, 2009), differences in practice mean that the distance from the noise and the volume of the noise being made are not always standardised, and when assessors are using different venues the echo and resonance of the sound may change, potentially impacting the test outcome. The Greater Manchester Police Dog selection tests (observed in January 2014) include a loud metallic bang while the dog is investigating another part of the room, following the bang they will call the dog over towards the metallic object and assess if they are willing to investigate, how long it takes them to approach the object, as well as any body language suggesting fearful behaviour such as ears back, flattening, tail tucked, some will also assess recovery time (dog to resume behaviour seen prior to noise), as suggested by Batt *et al*, 2008 recovery rate is potentially of more use than the startle response itself. As well as this assessors may perform some testing at height to check confidence by placing the dog on a platform as observed during puppy selectin tests at GMP in January 2014, tunnel or ramp as described in Asher *et al*, 2013 to assess confidence.

The dog's response to a stranger is often assessed in two ways, with a non-threatening and threatening person (Slabbert & Odendaal, 1999. Svartberg, 2002) unknown to the dog to ensure the dog is sufficiently sociable and confident in the face of threat. Variations of the test devised from the scientific literature include measuring the dogs response to human approach at different speeds (DeMeester *et al*, 2008), defence of self and handler from threat (Ruefenacht *et al*, 2002), and aggression (Svartberg & Forkman 2002). This portion of the test is often used more in protection than detection dog selection. Most commonly the approach of a stranger in dark clothing is used in practice, and often includes the stranger shouting or banging objects (not dissimilar to the 'ghosts' test reported in Svartberg & Forkman, 2002 used white sheets to cover the strangers while banging and moving erratically around the dog) while assessing the dogs body language to look for responses that would indicate levels of confidence in the dog.

Sociability with people is also tested for during selection, in practical field assessments it is often examined during testing based on observations of how the dog interacts and responds to touch from the tester and/or a stranger. This has come from literature examining dog behaviour when greeting people (Svartberg, 2002) and dog interactions with humans in a friendly situation (Wilsson & Sundgren, 1998. Ruefenacht *et al*, 2002. Svartberg & Forkman, 2002. Svartberg 2005. Svartberg *et al*, 2005. Svobodova et al, 2008). Practical field tests for working dog selection are often briefer and less controlled than tests conducted within the literature aimed to determine factors such as attachment in dogs (Duffy & Serpell, 2012), or those measuring factors such as the latency of contact the dog makes with the person, and the latency to initiate that contact (Batt *et al*, 2008). Testers may also incorporate a level of restraint handling in testing, mainly to see if the dog resists being held or show signs of distress, a test adapted from Asher et al, 2013 assistance dog selection testing where dogs are assessed while being stroked, restrained and their behaviour following such exposures to test sociability.

Selection tests will also often involve an element of play interaction with a toy, either through tugging games or a ball depending on the working role anticipated for the dog (tug and chase for protection, ball for search), to assess if the dog has sufficient play drive which is viewed as important for training a working dog. Some tests will assess retrieval (Wilsson & Sundgren, 1998. Svobodova *et al*, 2008. Asher *et al,* 2013), chase (Svartberg, 2002. Svartberg 2005. Svartberg *et al*, 2005. Batt *et al*, 2008. Gosling *et al,* 2009. Duffy & Serpell, 2012), tugging/ragging (Svartberg, 2002. Svartberg

& Forkman 2002. Svartberg, 2005. Svobodova *et al,* 2008. Wilsson & Sundgren, 1998), and ball play (Wilsson & Sundgren, 1998).

Desire for work is assessed usually through an obstacle course (Slabbert & Odendaal 1999. Svobodova *et al*, 2008) or working task such as obedience (Svartberg, 2002. Duffy & Serpell, 2012) or search (gosling et al 2009), to assess willingness to work. In a practical field setting it has been observed in terms of a detour task to assess problem solving and persistence, although it varies across practice this method combines overcoming obstacles with a simple search task.

Using the results of these tests assessors will then score the dog on a number of behaviours shown in the various contexts of the tests to identify traits of interest, usually along the lines of confidence, play drive, willingness to work and fearfulness before making a decision on if the dog meets their criteria in terms of these traits for selection. The tests being implemented in the field are generally less controlled than those they are based on out of the literature, and testers often develop their own methods of scoring the dogs, which is usually centred around their own experiences of the dogs they have encountered which have been successful in work, or they are scoring dogs against others in the same selection cohort. In the field it has been observed that if a set number of dogs need to be selected then the standards may change to reflect the group of dogs testers have to choose from (observations of GMP puppy testing January 2014, the whole cohort could have been selected based on the criteria, but finances only allowed a small selection of dogs to be taken forward for training).

By using the current literature to inform selection tests, testers are trying to ensure that only suitable dogs are being selected for work. However, not all dogs that are selected end up being certified, and not all dogs certified remain in service for the entirety of their working life. This leads to the suggestion that other factors may be impacting on working dog withdrawal.

1.5.2 Factors Affecting Performance

Working dog performance can be affected by a number of factors including their physical fitness levels (Gazit and Terkel, 2003), age (Tiffoli & Rolfe (2006), their relationship with their handler (Lefebvre *et al*, 2007) training contexts (Gazit *et al*, 2005), environmental challenges (Haverbeke *et al*, 2008) (including factors such as depth of burial of an explosive device for detection dogs (Diverio *et al*, 2016), as well as climate factors such as temperature extremes (Tiffoli & Roilfe, 2006) and wind direction (Diverio *et al*, 2016). If a dog is unable to cope with such factors, it could induce stress which may then have a negative impact on welfare and hinder performance in the field (Diverio *et al*, 2016). Dogs respond to stressful and distracting situations differently (Horvath *et al*, 2007), and this is thought to be related to individual differences in coping styles (Koolhaas *et al*, 1999). While performance is affected by a number of factors, temperament is thought to underpin how dogs cope with these factors.

The testing of potential working dogs for their aptitude to work prior to training is an area of ongoing scientific research (Sinn *et al*, 2010), but ability does not necessarily predict performance and resilience under challenging conditions. In some cases, 48% of dogs recruited for the military training process fail their certification test in the USA (Foyer *et al,* 2015). While every effort is made to emulate field situations in training, because of the extreme nature of some of the environments dogs are put to work in, despite the fact that dogs are required to pass a certification test before they are signed off to operate within the field, a proportion of dogs end up being withdrawn from service due to a failure to perform in an operational setting. Failure to perform at an adequate level has the potential to endanger the general public and even national security. As many as 9% of military working dogs fail to certify and some reports state up to 19% of military working dogs in the USA die or are euthanized because of behavioural problems (Moore *et al*, 2001). With behavioural problems being reportedly the cause of withdrawal from active service in 82% of cases in the USA (Evans *et al*, 2007), while the current system works for the majority of dogs there is a need to re-evaluate methods for selection and training to maximise the number of dogs that certify and reduce the number of dogs withdrawn from service due to behavioural reasons.

1.5.3 The Role of Temperament in Selecting Working Dogs

Temperament has a role in working dog success in many areas including in terms of engagement and focus which are key performance criteria for working dogs related to resilience and distractibility, especially in an operational context. Distractibility has been shown to be an important trait in predicating certification of detection dogs (Maejima *et al*, 2007) and guide dogs (Arata *et al*, 2010). Distraction may occur in the presence of either exciting or fear eliciting stimuli. In other words working dogs need to concentrate and work in the presence of temptation as well as threat. Predicting resilience and stress is of importance for selecting suitable working dogs (Rooney *et al*, 2016). Resilience covers factors such as response to stressful exposures, performance under stress, and time taken to recover following stressful experiences (Southwick *et al*, 2005). Although resilience per se does not appear to have been studied directly in working dogs, there is evidence that the related concept of confidence predicts certification; dogs showing confidence in behaviour tests may be demonstrating an ability to cope with potentially distracting or stressful situations (Slabbert & Odendaal, 1999) and so are more resilient. It is important for working dogs to be resilient to ensure that they can continue to work effectively despite being exposed to challenging and potentially aversive environments, however, this is clearly a multifactorial feature which has no simple correlate.

1.6 Measures of temperament:

Temperament in humans has been assessed using a combination of trait constructs, behavioural tests, questionnaires and social cognitive measures, while in animals, standardised behavioural tests, psychometric scales and subjective ratings are more typically used. A lack of consensus over definitions and terminology relating to temperament and its various component traits is problematic in this field (Morris, 2000. Rothbart *et al*, 2000. Judge & Ilies, 2002), but despite the inconsistent terminology (Svartberg *et al*, 2003) there is evidence of convergent validity of measures of temperament in a number of species (Reale *et al*, 2000. Rothbart *et al*, 2000. Seaman *et al*, 2002. Deiderich *et al*, 2006. Reale et al, 2007. Sinn *et al*, 2010). In dogs, temperament tests ideally need to demonstrate aspects of reliability and validity to be useful (Stamps & Groothuis, 2010). For the purposes of this thesis reliability and validity are defined below.

1. Reliability: The degree to which the result of a measurement, calculation, or specification can be depended on to be accurate.
    a. Inter-rater reliability - Level of agreement between raters
    b. Intra-rater reliability - Level of agreement between the same rater on repeated tests
    c. Test-retest reliability – Test consistency over time
2. Validity: The quality of being logically or factually sound
    a. Construct Validity – The extent to which a construct measure what is theoretically should
    b. Convergent Validity – the extent to which theoretically related variable are factually related
    c. Discriminant Validity – the extent to which theoretically unrelated variables are factually not related
    d. Predictive Validity – the extent to which measures in test predict measures in the field and performance in other related tests

The more measures of reliability and validity a test can provide, the more accurately it can be interpreted, so ideally a number of measures would be taken and examined in a test to establish reliability and validity.

Tests may not always be predictive in contexts outside of the test context, which raises the issue of test construct validity, and demonstrates a need to ensure that the trait being measured is indeed a trait (behavioural tendency), rather than simply a specific behavioural response to a specific situation. For example, a behavioural response to a gun shot or loud noise could be a startle response which is seen in a number of contexts, this however is a behavioural response rather than a trait. A dog that startles on hearing a loud noise is not necessarily fearful or lacking in confidence, a dog that hears a loud noise and then becomes anxious and attempts fight or flight strategies, and adopts this approach in other potentially threatening situations, could be ascribed the trait fearful, although traits need to be based on a number of behavioural responses across a number of contexts (Sinn et al 2010).

1.6.1 Temperament Assessment

Psychometrics is a branch of psychology examining the measurement of knowledge, abilities, attitudes, and personality/temperament traits (Rust *et al*, 2014). Physiological measures can be used to assess correlation (convergent validity) with what is observed in behaviour tests or psychometric profiles derived from carefully constructed questionnaires. This sort of combination of test methods can be used to establish the validity of temperament constructs in animals and has been achieved to an extent in equines for very simple traits (Momozawa *et al*, 2003), and in dogs in terms of trait impulsivity (Wright *et al*, 2011. Wright *et al*, 2012. Reimer *et al*, 2014).

In dogs, perhaps one of the most important applications of temperament research is in selecting dogs for working roles (Wilsson, 1996). Working dog groups need methods to detect behaviours that could potentially be problematic or beneficial in work early on in the selection process to avoid recruiting dogs that would be unable to cope in the working environment and not work effectively as a consequence of this. Ability to perform a task is insufficient to predict performance in the field as additionally dogs need to cope with the working environment and some working dogs are more successful in this than others. This variation in performance is thought to be due, at least in part, to individual differences in temperament, but it is unclear which temperament traits are most strongly related to long term working dog success (Sin *et al*, 2010).

There has been extensive interest in temperament traits that can be used to predict certification in working dogs, but this is only the beginning of a working career and there appears to be a lack of focus on traits that contribute to a long and successful working life. Additionally, there is a lack of consensus regarding the definition of a “successful” working dog (Foyer, 2015), although preferred traits have been identified (e.g. for search dogs, Rooney *et al*, 2004). Lifetime performance is probably one of the better measures available but selection tests are not always predictive of good performance in the field (Sin *et al*, 2010). Police and military working dogs are faced with challenging conditions which they need to be robust enough to cope with and still work. The identification of individuals which are likely to cope (as well as be capable of the tasks asked) is an important factor in the selection process so it is necessary to have available tests predictive of this to assist with selection of working dogs.

1.6.2 Temperament and Working Dogs

Current research has made solid contributions to working dog selection processes and aided understanding that behavioural characteristics are often responsible for failure rather than physical or sensory attributes (Foyer *et al*, 2014. Maejima *et al*, 2007. Rooney *et al*, 2007). Working dog selection is an area of ongoing research where methods of assessment of working dogs are constantly being developed and revised to prevent withdrawal in the field due to behaviour (Sinn *et al*, 2010).

Fear is recognized as an important trait in terms of working dog performance (Rooney *et al*, 2016) and has been shown to be a heritable trait (e.g. Ruefenacht *et al*, 2002. Ilska *et al*, 2017). Fear based behaviors in a working dog role result in dogs not completing their training (Murphy, 1995. Batt *et al*, 2008), and being withdrawn from work due to behavioral reasons (Caron-Lormier *et al*, 2016). While the importance of fear in terms of working dog withdrawal is recognised as a concern, selection tests aim to assess fear related behaviour to minimise this and are seemingly effective as a number of dogs selected do work effectively for the duration of their service, but a number of dogs still fail or are withdrawn. It is likely that is other factors are potentially also of importance in contributing to working dog withdrawal and failure. Sensitivity to negative stimuli (a contributing factor to fear based behaviour) is also a feature of resilience (ability to cope with the negatives), which is another factor affecting general (rather than task specific) performance and ability to cope. Resilience has been less directly assessed in working dogs, although it could be argued that resilience has been examined in terms of confidence which predicts certification: with greater confidence in behaviour tests indicating a better ability to cope with potentially stressful situations (Slabbert & Odendaal, 1999). Generally traits such as confidence and desire for work have been shown to be of importance in terms of working dog certification (e.g. Maejima *et al*, 2007), although working dog withdrawal from service is still an issue. With up to 70% of police dogs bred for work being unsuitable in South Africa (Slabbert *et al*, 1999), around 70% of detection dogs in Japan failing to complete training (Maejima, 2007), and just over 50% of those bred for work failing to certify in the Swedish Armed Forces (Foyer *et al*, 2016), there is clearly an international problem in selecting working dogs. While existing selection procedures ensure a number of dogs are suitable for work, there are likely other factors contributing to withdrawal that are currently not being identified.

1.6.3 Resilience and Distractibility

To further understanding of which dogs are likely to cope in the operational domain it may be useful to consider further ways to predict resilience and distractibility in addition to current traits regularly assessed in working dogs. These concepts tap in to mechanisms that logically would be of importance in an operational context. Resilience being an ability to cope under stress and distractibility being able to be re-engaged by the handler back into work when they lose focus.

Distractibility has been shown to predict certification of detection dogs (Maejima *et al*, 2007) and guide dogs assessed as being easily distracted on their trainer completed questionnaires were more likely to be disqualified from service (Arata *et al*, 2010), and this component of distraction has been linked back to puppy raiser questionnaire responses which suggests distractibility resulting in disqualification could be identified from an early age (Kobayashi *et al*. 2013). Also in guide dogs, questionnaire measured tendency to pull excessively hard on the leash (which could be related to frustration at not being able to investigate distractions), was linked to disqualification from service (Duffy & Serpell, 2012). This suggests distractibility as an important component when it comes to working success.

Working dogs need to be able to focus on work despite challenging conditions which may reduce focus to task. In dogs, the constructs of positive and negative affect have been used to describe behavioural tendencies linked to sensitivity to rewards and punishers in the environment (Sheppard & Mills, 2002). How sensitive a dog is to positives or negatives within the environment potentially relates to distractibility (as attention is being drawn towards these qualities in the environment, rather than directed towards the task they are required to perform). This distractibility may be thought of as having at least two emotional forms, one related to positive stimuli (how salient positive environmental stimuli are to the dog) and one related to negative stimuli (how salient the dog finds aversive qualities within the environment), these have been examined in terms of core affect within the literature (Taylor & mills, 2006. Russel, 2009).

1.6.4 Core Affect:

Core affect is central to emotion, it is feeling good or bad, it is not a reflective process but it can be altered by events, whether in the past, present or future (Russel, 2009). Reflective information that is cognitively processed has an effect on core affect but as this information fades from consciousness so its impact on core affect also dwindles (Russel, 2009).While reflective information will have an effect on core affect, core affect itself will impact on behavioural responses across a range of situations and contexts. For example, individuals that are more sensitive to factors eliciting the ‘bad feeling’ (negative activation) are likely to be easily distracted by potentially threatening stimuli in the environment. While those more sensitive to factors eliciting the ‘good feeling’ (positive activation) are likely to be more resilient and more easily have their attention redirected by positive incentives. The two are not mutually exclusive and well balanced individuals are likely to experience low to mild levels of negative activation (so they are aware of potentially threatening stimuli but can cope and recover from stressful exposures), and relatively high levels of positive activation (enabling them to redirect from stressful exposures into more useful and constructive activities) (Russel, 2009).

Distractibility and resilience will depend partly on the elements of core affect in terms of sensitivity to rewards (positive distractions - temptations) and aversives (negative distractions – anxieties) and a psychometric instrument for the assessment of sensitivity to rewards and aversives in dogs has been developed (Positive and Negative Activation Scale -PANAS, Shepherd and Mills 2002). This has been found to be robust in animals as young as 10 weeks of age (Sheppard and Mills 2002) and have predictive validity in relation to associated problem behaviour (fears and problems of excessive seeking behaviour (attempts to obtain desires)) (Mills *et al*., unpublished data) although to date there are no existing behavioural measurements for the PANAS, meaning prior knowledge of the dog across a range of contexts is required to complete it.

Behavioural expressions of both sensitivity to positives and sensitivity to negatives are likely moderated by trait impulsivity. Trait impulsivity impacts the rate at which decisions are made and actions are taken (Wright *et al*, 2011). In working dogs there is perhaps value to having dogs that make quick and unambiguous decisions, this may be a feature of impulsivity, but if the reaction to negative stimuli is impulsive (without proper evaluation) this may be perceived as an undesirable trait.

1.6.5 Impulsivity:

Impulsivity is a multifactorial construct based around the notion of forethought to the consequences of actions (Peremans *et al*, 2002). Impulsive individuals will show behaviour characterized by little or no thought or consideration of the consequences (Evenden, 1999). Impulsivity in the dog can be assessed psychometrically using the Dog Impulsivity Assessment Scale (DIAS- Wright *et al* 2011) which has been validated with both behavioural (delayed reward choice task) and physiological (urinary metabolites of serotonin and dopamine) measures in dogs (Wright *et al*, 2012). Impulsivity in the DIAS consists of 3 elements as well as overall impulsivity; the constituent elements are “Responsiveness”, “Aggression & Response to Novelty”, and “Behavioural Regulation”. All of which may play a role in decision making, attentional focus, and ability to cope, which are likely to impact working dog success. Additionally, when assessed in this way, these measures of impulsivity have been shown to be stable for up to 6 years (Reimer et al, 2014). Impulsivity is also potentially relevant to the assessment of distractibility since it describes the extent to which an individual evaluates the consequences of their behaviour (Peremans *et al*, 2002), and is expressed in a range of behaviours linked with inhibitory control, trainability and focus (Wright *et al*, 2012). More impulsive individuals may be expected to be more easily distracted, if all other things are equal (Peremans *et al*, 2002).

1.6.7 Summary

There exists a need for further tools to determine qualities linked to withdrawal in working dogs as an addition to current measures to improve the predictions of such tests already in use. The process of collection of information to select working dogs needs to be quick, simple and effective for use in the field as time to assess each dog is limited and where dogs are assessed is often varied depending on where dogs are being sourced from. Resilience and distractibility are important components for working dog success as dogs need to be able to cope in varied environments and not be distracted by either excitatory or fear eliciting stimuli. While questionnaires such as the C-BARQ have proved useful in aiding the selection of working dogs (Foyer *et al*, 2014), other measures such as the PANAS and DIAS have potential value. While these have been used in pet dogs to assess and identify links to behavioural problems, applications to working dog populations have not been attempted with these tools. In this thesis it is argued that the tendencies of resilience and distractibility relate to the constructs of positive and negative activation alongside impulsivity, which already have instruments developed for their assessment in dogs. Sensitivity to positive (excitatory) and negative (fear eliciting) qualities in the environment can be assessed using the Positive and Negative Activation Scale (PANAS) (Shepperd & Mills, 2002), and impulsivity can be assessed using the Dog Impulsivity Assessment Scale (DIAS). Together these scales provide a framework for assessing differences between successful and failing dogs in the field. As the field of working dog research expands it would be useful to understand if these tools can be applied to the working dog sector, in a similar way to the C-BARQ to potentially provide more information on selection of other traits of relevance and to build on the existing body of research. This could provide another tool for assessing working dogs and predicting individual differences in performance.

1.7 Aims & Objectives

1.7.1 Aim:

This project aimed to identify and develop additional tools to aid to the selection of successful working dogs associated with the police and military.

1.7.2 Objectives:

1. Explore the relationship between individual arousal and performance

An inevitable product of individual differences is the possibility that individuals may have their own individual preferred level of arousal at which they function optimally, rather than there being a task specific optimal level of arousal for its execution. This can be explored by examining the relationship between arousal and performance in a working dog task.

2. Review the current literature on behaviour tests used in the selection of working dogs to bring together current knowledge on working dog selection tests and identify gaps in the literature.

There is a need to understand behaviours that are measured in current working dog selection tests, and how performance in such tests translates to performance in the field. This can be achieved objectively through the use of a review of literature relating to working dog selection tests.

3. Develop and validate a new behavioural test to assess impulsivity in dogs that is faster and more inclusive than the existing behavioural test.

High levels of trait impulsivity are linked to behavioural problems in dogs, because of this and the traits link to distractibility it may be important in terms of working dog withdrawal. Working dogs need to be assessed quickly and tests need to be simple enough for the majority of dogs to complete as time is often limited when working dogs are being purchased so testing needs to be simple and quick to implement. Currently the existing test (Wright *et al*, 2012) is time consuming and only a small number of dogs are able to reach criteria for evaluation with the test. These are major limitations to the practical use of the currently available test, a new test will be developed to behaviourally assess impulsivity without such limitations.

4. Assess working dog success versus withdrawal in terms of temperament traits hypothesised to be of value in a working dog population.

As mentioned above, impulsivity may relate to working dog withdrawal, but the extent to, and way in which this is the case remains unknown. In order to be able to predict working dog success or failure using these tools, there is a need to quantify what makes a successful working dog. This will be achieved through the profiling of successful and unsuccessful working dogs using already-validated instruments not only for impulsivity but other theoretically important traits as well (core affect). Temperament traits related to impulsivity and positive and negative activation will be assessed.

5. Develop behaviour tests that can distinguish between dogs matching a successful and withdrawn working dog profile.

In order to effectively and efficiently assess working dogs for relevant tendencies, simple behavioural tests need to be developed to enable assessment of the constituent features of the traits found to be of relevance in success versus failure as a working dog. Unlike psychometric tools these do not depend on prior knowledge of the subject or reporter honesty in relation to the preferred phenotype. Behavioural tests for working dogs need to be simple and effective so they can be applied to a large number of dogs relatively quickly.

Chapter 2: Arousal & Performance

Synopsis:

Arousal results in the non-specific activation of sensory systems, increased heart rate and a general state of alertness. This state can be positive or negative in valence and impacts on task performance, as an individual is more likely to be distracted if the level of arousal is not conducive to the task at hand. Based on this notion, two laws of arousal and performance are widely cited, the Yerkes Dodson law and the Individual Zones of Optimal Functioning (IZOF) law. The first of these suggests that arousal levels required for optimal performance are task specific, and arousal levels above or below this result in a drop in performance. The second law, IZOF, suggests that individuals have their own personal levels of optimal arousal for a task. These laws are not necessarily mutually exclusive. The present study sought to examine the link between arousal and performance in a scent detection task in pet dogs, with the aim to identify if there was an optimal level of arousal for the task, individual optimal levels of arousal for task performance, or if performance was related to individual differences based on temperament. The study recruited 20 pet dogs trained in the detection of catnip (as their target scent), profiled their temperament using the Positive and Negative Activation Scale (PANAS) and Dog Impulsivity Assessment Scale (DIAS), and examined their performance in a scent detection task under a range of levels of arousal (manipulated through different activities prior to searching) while recording their heart rate variability (HRV) to be used as a physiological measure of arousal. The results of the study revealed that the HRV measurement methods used were unsuitable for determining level of arousal. When activity related conditions were used to infer arousal levels, no relationships were found between condition and performance. An element of the temperament traits profiled (Negative Activation) using PANAS was correlated with average search score, with individuals scoring higher in the trait receiving higher search scores. DIAS Overall Questionnaire Score showed near significant positive correlation with average search score as well. It was uncertain if the same conditions elicited the same level of arousal in the dogs and overall it was difficult to determine the impact of arousal on performance, however measures of temperament appeared to be linked to performance so this line of enquiry was pursued further.

2.1 Introduction:

2.1.1 Stress & Arousal

Stress is primarily a physical response resulting in the release of compounds such as cortisol and adrenaline preparing the body for 'fight or flight' (Lindau *et al*, 2016). Stress has been described as the result of an imbalance between the demands being placed on an individual and their ability to meet these demands (Gerber *et al*, 2008). In humans short bursts of low level stress have been shown to actually increase memory recall, while high levels of stress or long lasting stress has a negative impact on memory function, this affect has also been observed in animals where low levels of stress have been found to be beneficial for spatial learning but higher levels of stress decrease performance (Lindau *et al*, 2016).

While stress has been shown to be of importance in terms of performance, arousal covers a broader range of emotional states so is the focus of this study. For the purposes of this research arousal will be referred to throughout as defined here; arousal is a function of awareness and impacts on direction of attention to guide appropriate future action (Gerber *et al*, 2008). It includes factors associated with positive and negative states of emotion such as stress (Gerber *et al*, 2008) but encompasses a broad spectrum of emotional states such as excitement and anxiety (Barret & Russell, 1998).

2.1.2 Arousal and performance

As discussed previously in Chapter 1, performance is likely to be affected by a number of factors ranging from environmental challenges to individual differences in behaviour. Another factor thought to impact on performance is the level of arousal experienced prior to participating in a task.

There are two main theories in terms of how arousal can impact on performance, one is based on the notion that certain tasks require certain levels of arousal to be performed optimally (Yerkes Dodson Law (Yerkes & Dodson, 1908)), the second is that individuals have different levels of arousal within which they function at their best (Individual Zones of Optimal Functioning-IZOF (Hagvet and Hanin, 2007)).

2.1.3 Yerkes Dodson Law of Arousal and Performance

The Yerkes Dodson law proposed in 1908 (Yerkes & Dodson, 1908) suggests that too low or too high arousal impair the amount of information that can be attended to, therefore limiting performance.

Set levels of arousal can enhance performance through promoting learning or change although too much or too little arousal has a negative impact on performance (Anderson *et al*, 1989). The Yerkes Dodson law proposes that optimal levels of arousal for performance are task specific, with cognitive tasks requiring lower levels of arousal for optimal performance, while endurance or persistence tasks require higher levels of arousal for optimal performance (Yerkes & Dodson, 1908). See figure 2.1 below.

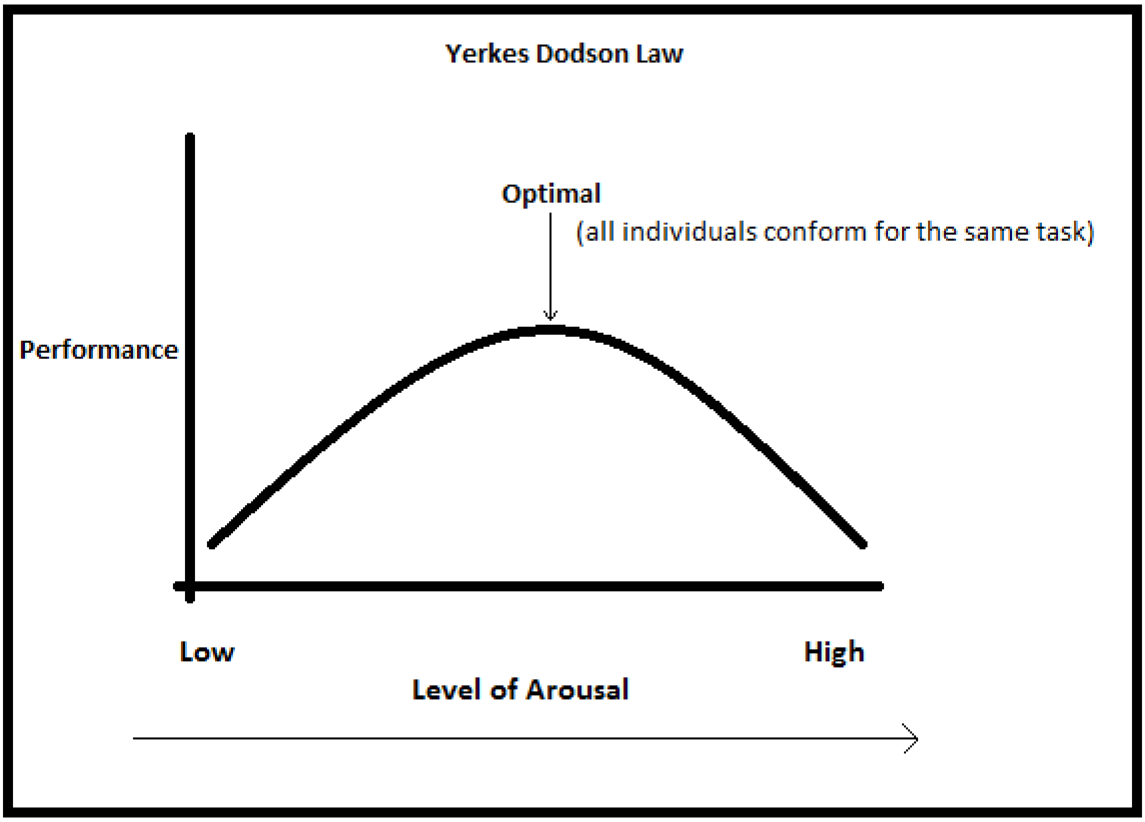


Figure 2.1 Representation of the Yerkes Dodson Law of level of arousal for optimal task performance where the x-axis represents the level of arousal in terms of activation from low (left) to high (right) and the y axis represents performance from low (poor) to high (optimum).

However, the Yerkes Dodson law doesn't account for individual differences which may impact on how arousal affects performance or assumes that these effects are minimal; an alternative perspective is to suggest that optimal arousal levels are not largely related to a task, but rather more about the individual.

2.1.4 Individual Zone of Optimal Functioning

The concept of the Individual Zones of Optimal Functioning or IZOF hypothesis is developed from a sports science concept which suggests that an athlete performs at their best while experiencing an ideal range of anxiety (Hagvet and Hanin, 2007). It is suggested that this changes for different sports (so does consider Yerkes Dodson law) and that each individual will have their own range in which they function at their best (Hagvet and Hanin, 2007). Thus it is the individual who largely determines the optimal level of arousal rather than the task. See figure 2.2 below.

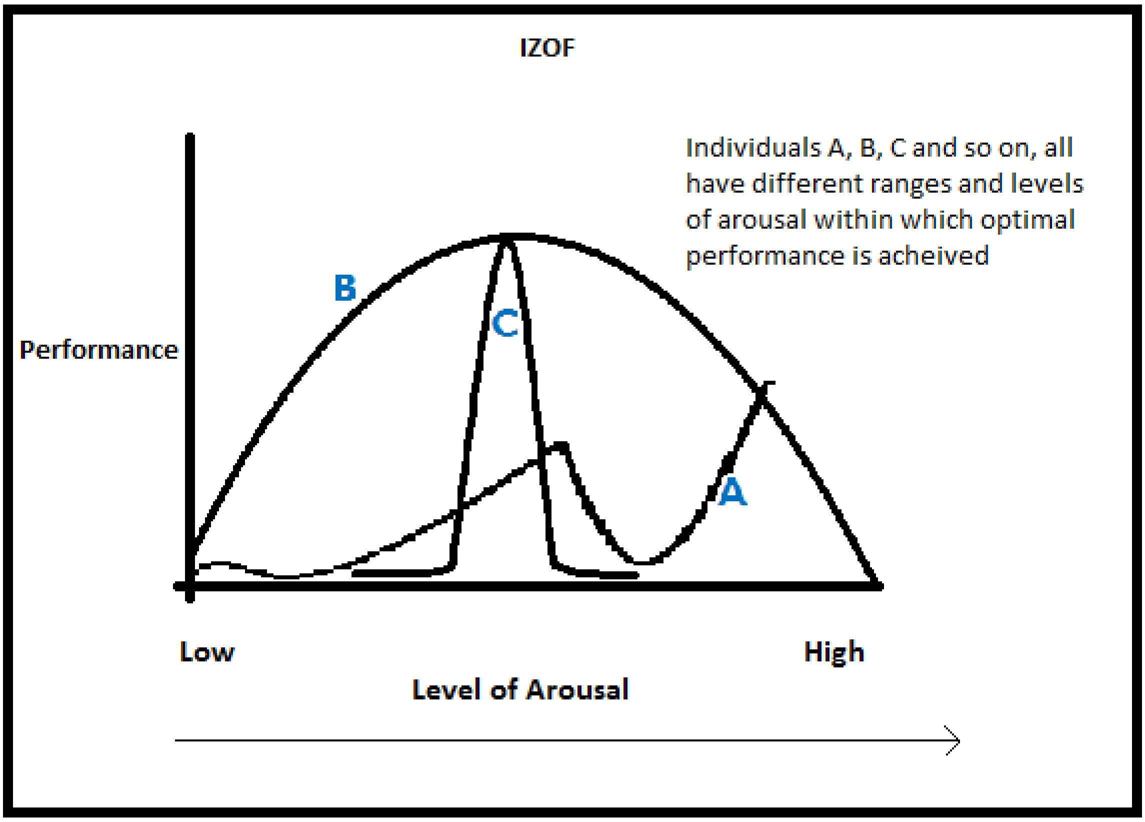


Figure 2.2 Representation of the Individual Zone of Optimal Functioning (IZOF) theory of individual variation in level of arousal for optimal performance in a single task. Where the x-axis represents the level of arousal in terms of activation from low (left) to high (right) and the y axis represents performance from low (poor) to high (optimum). A, B and C represent different individuals taking part in the same task experiencing optimal performance at differing levels of arousal to one another.

It has been suggested that somatic measures of anxiety such as cortisol or heart rate variability are more reliable than cognitive measures such as self-report (Robazza et al, 2012), although cortisol (which is actually a more general measure of arousal, rather than a specific somatic measure of anxiety) can affect cognitive functions by influencing the way information is processed and inhibiting information processing in some cases (Huang and Hung, 2010). This is useful information as in non-human animals self-report is not possible, creating a reliance on somatic measures which are preferable to inferring levels of anxiety or arousal, which is largely limited by interpretation. It is suggested that these measures are task specific and will change dependant on the demands of the task being performed as in line with Yerkes Dodson law (Huang and Hung, 2010). For example in Rugby increased cortisol levels are linked to improved performance, while the opposite has been found for dancers (Huang and Hung, 2010). Salivary cortisol has successfully been used to identify

IZOF in competitive shooting showing that pre competition somatic anxiety (as it is described by the authors) within a set range resulted in better performance on an individual basis (Robazza et al, 2012). Thus while there seems to be a general association with task, IZOF suggests it is more important to attend to variation between individuals.

Some reports suggest that mental training to enable athletes to enter their optimal zone can enhance their performance and interventions to manipulate emotions (to influence the zone an athlete is in before performing) have an impact on performance (Hagvet and Hanin, 2007). Interventions to influence the emotional state of athletes to map on to IZOF have demonstrated some success in improving performance (Robazza et al, 2004), and it has been identified that athletes perform better when working inside their optimal zone (Salminen et al, 1995). However, these measures focus on reports from athletes on their emotional state before performance and reflecting on their awareness and acceptance of this state (Salminen et al, 1995. Hagvet and Hanin, 2007), as this is not an option in dogs, other non-invasive methods such as heart rate variability (a measure of sympathetic-parasympathetic tone) can be used to infer arousal alongside pre performance interventions designed to influence emotional state.

The laws of IZOF and Yerkes Dodson are not necessarily mutually exclusive. It could be that certain tasks do require lower levels of arousal as stated in Yerkes Dodson law, but within this individuals may perform better within different ranges of low arousal, as per IZOF law, or may have a difference in the valance of their arousal, as arousal can either be positive or negative, which likely also impacts performance.

2.1.5 Arousal and Valence

Research into level of arousal and valence in humans has resulted in the development of models illustrating the relationship between the two factors. Arousal results in emotional states which are generated through a combination of the level of arousal (it's intensity or activation) from low (deactivation) to high (activation), and it's valence (whether it is a pleasant (positive) sensation or an unpleasant (negative) sensation). Three main models exist in human psychology literature to explain the relationship between arousal and valence.

1. The Bipolar distribution model suggests that activation increases in valence from negative to positive (Figure 2.3, left hand side).
2. The U shaped distribution suggests that regardless of whether the valence is positive or negative, it increases from most neutral to most intense ((Winston *et al*, 2003, 2005; Cunningham, *et al* 2004 (Figure 2.3, right hand side)).

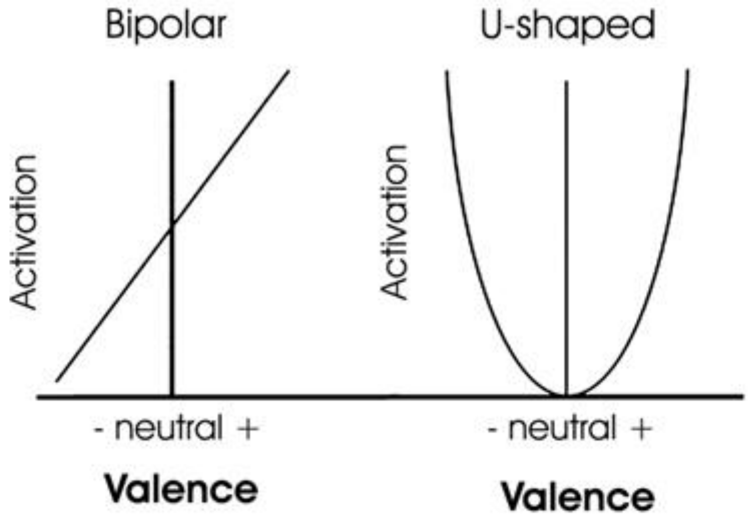


Figure 2.3 Depictions of the Bipolar model (pictured on the left) of valence and arousal compared to the U shaped model (pictured on the right) of valence and arousal. Diagram developed and adapted from Barrett & Russell, 1998.

3. In the independent model of arousal and valence, arousal exists along two independent scales, one of level (low to high (also referred to as inactive to active)) and one of valence (ranging from positive to negative (also referred to as approach to avoidance)) (Gerber *et al*, 2008. Barrett & Russell, 1998). See Figure 2.4 below.

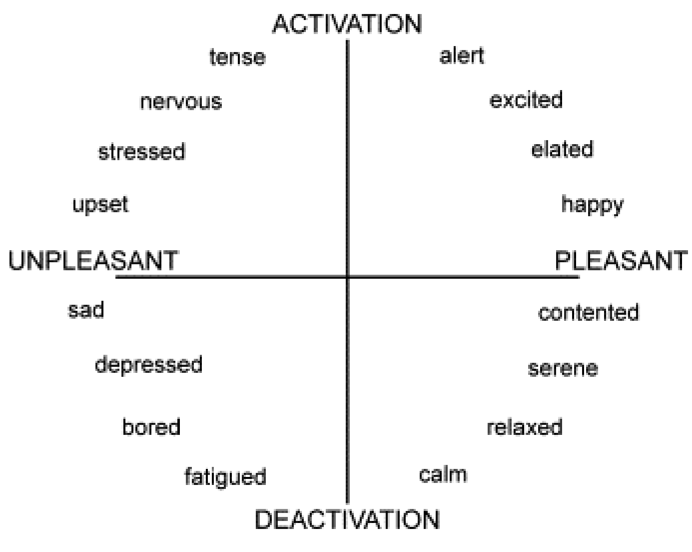


Fig. 2.4. The diagram depicts each emotional state along continuous dimensions of arousal (*y*-axis) and valence (*x*-axis) (Barrett & Russell, 1998).

The Independent model fits more within the IZOF principle. The IZOF principle conceptualises that level of arousal and valence both have an impact on performance and divides valance into pleasant (positive) and unpleasant (negative), while dividing level of arousal (activation) into helpful (functional) and harmful (dysfunctional) (Lautenbach et al, 2014). From these classifications 4 states are derived. P+ is a positive functional state, N+ is a negative functional state, P- is a positive dysfunctional state and N- is a negative dysfunctional state (see figure 2.5). These are thought to cover the most important potential states and their effect on performance in relation to IZOF (Lautenbach et al, 2014). People function optimally in either of the functional states, whether this is within a positive or negative functional state depends on the individual (Lautenbach et al, 2014), as does the extent to which each state is expressed or experienced by the individual. It acknowledges that functional levels of arousal for individuals can differ, as it does not simply state that high arousal is functional, and low dysfunctional, while that could be the case for one individual, another could have the reserve as applicable.

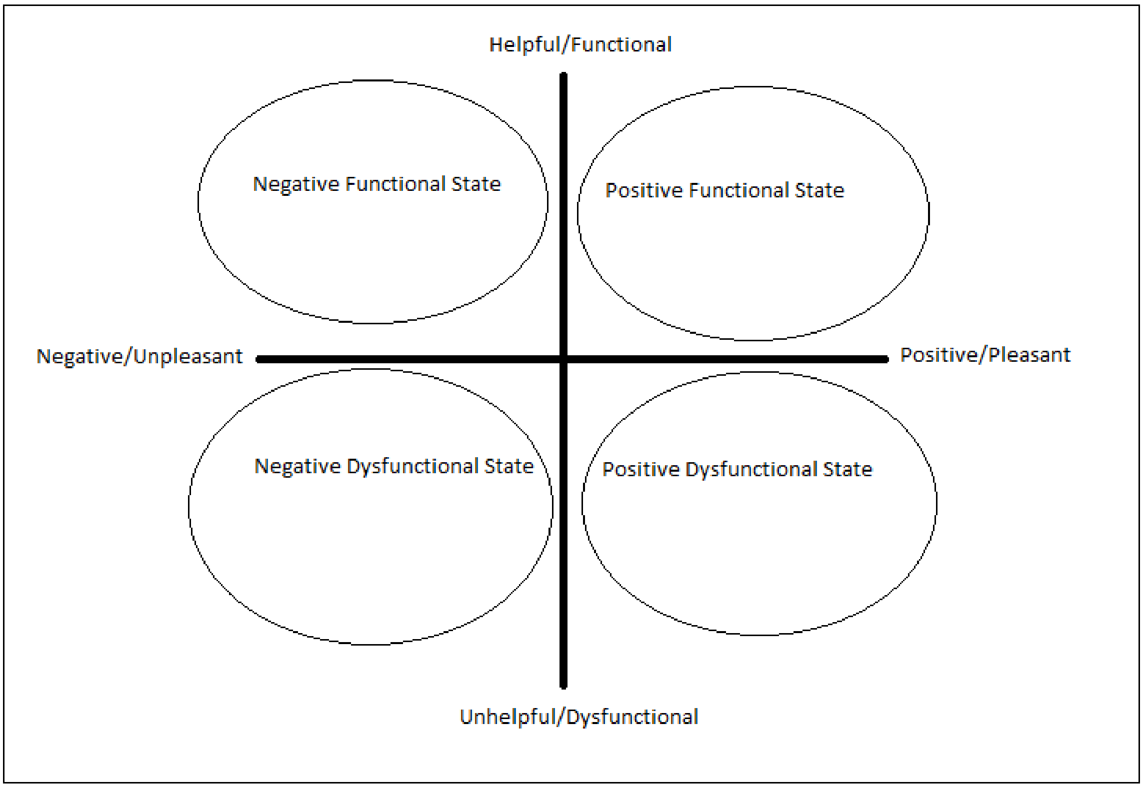


Figure 2.5 Model of the relationship between emotional valence and arousal (in terms of arousal being functional or dysfunctional), and valence being positive or negative to illustrate performance states as conceptualized by IZOF.

This model and the literature on arousal and valence highlights that it is important to not only consider the physical activation of arousal, but also the emotional valence as well.

2.1.6 Physiological Measures of Arousal

Heart Rate is usually defined as the number of beats per minute (HR), and is often used to assess health. It is generally affected by factors such as fitness and will alter during exercise (Lane *et al*, 2009. Luft *et al*, 2009). Heart Rate Variability (HRV) is the variation in time intervals between heartbeats, and is thought to be more affected by emotional arousal, rather than physical state (Lane *et et al*, 2009. Luft *et al*, 2009). HRV is thought to be more sensitive to cognitive processes occurring within an activity (internal states). It is related to emotional regulation in humans (Lane *et al*, 2009. Folkman *et al*, 1986. Lautenbach *et al*, 2014) and is not as sensitive to movement as heart rate itself so it is possible to get good readings during activity, although the more energetic the movement the more errors there will be in the data. This can be rectified using error correction methods on the data, through methodology developed by Schoberl *et al*, 2015. SO it would be possible to measure during a working dog task.

2.1.7 Testing Search Ability

In dogs, search ability has been widely assessed to help identify suitable individuals for detection work (Rooney *et al*, 2007). Search tests involve locating one or more pre trained scents, and often consist of a free search (where the dog is left to search without handler involvement, except for the cue to search, and when the dog indicates), and a systematic or directed search (where the handler guides the dog through the search area to ensure all potential areas have been examined by the dog, as described in Rooney *et al*, 2007). The location of the item the dogs are searching for is standardised for all dogs, but changed on repeat searches for the same dog, to allow for comparisons between individuals in terms of their performance, which can be measured in terms of the speed of the find, or the number of finds successfully located (Rooney *et al*, 2007). While in working dogs search tests are used to check the dogs ability, individual factors affecting performance and the relationship between arousal and performance has not been examined to our knowledge.

2.1.8 Individual Differences and Performance

As discussed in Chapter 1, individual differences are likely to impact on how an individual performs and functions, so this is an important consideration as well as models of arousal. As previously discussed the PANAS and DIAS are tools that can be used to identify individual differences in terms of impulsivity and core affect. The PANAS may also be tapping into how individuals experience and are predisposed to valence in terms of arousal as it examines sensitivity to positive and negative qualities in the environment. These tools could be used to cautiously examine the role of temperament traits in performance if measures of HRV fail and/or the arousal scales are not related to performance.

2.1.9 Aim

The aim of this study was to identify conditions prior to working dog tasks that could influence performance, with a view to determining whether performance was related to task or individual arousal. In the event of a failure on measures of arousal, individual differences from profiling dogs using the PANAS and DIAS were to be examined to identify their impact on task performance.

2.1.10 Hypothesis

It is hypothesised that performance in terms of search score will be related best to either a given range of arousal in all dogs as determined by pre-test activity (in line with Yerkes-Dodson) or individual level of arousal as assessed by HRV (in line with IZOF).

If arousal measures failed, it was hypothesized that individual differences in the form of scores of component items taken from the PANAS and DIAS profiles would be linked to task performance.

2.2 Methods

2.2.1 Selection Criteria:

Dogs recruited for this study had all been trained on Talking Dogs Scent work competency level 1 (all the dogs had been trained how to detect a scent and give an active indication once they had found it, handlers had been trained on how to perform a free search and a directed search with their dog), using the same training techniques to detect the scent of catnip on small pieces of material.

To participate in the study the dogs were screened to ensure that they had no history of guarding food or toys from people, were described by their owners as being comfortable being handled (sensitively) by their owner and/or strangers (for fitting a heart rate monitor (HRM)). The owners were required to consent to their searches being filmed for analysis.

2.2.2 Recruitment:

The owners of 24 pet dogs meeting the inclusion criteria attended a pre-assessment session with the opportunity to participate in research looking at factors that could affect their dog's performance. See Figure 2.6 for information on age, sex and breed of the dogs recruited.

Age (years)	Sex	Breed
1	M	Border Collie
2	F	Cocker Spaniel
3	M	English Springer Spaniel
5	F	Crossbreed
3.5	M	German Pointer
4	F	Labrador
5	M	Terrier X Pug
2	M	Miniature Schnauzer
4	M	Labradoodle
5	F	Cocker Spaniel
6	F	Cocker Spaniel
3	F	Cocker Spaniel
6	F	Border Collie
4	M	Border Collie
6	F	Border Collie
5	F	Labrador
7	M	Collie X Kelpie
2.5	M	Cockapoo
11	F	Cocker Spaniel
2	F	Jack Russel Terrier X Staffordshire Bull Terrier

Figure 2.6 Demographic information on the dogs that participated in the study.

All training and testing was carried out at the University of Lincoln Riseholme Park Campus Training Barn, LN2 2LG.

Owners were asked to give consent in writing for their dog to take part in the project and for their searches to be filmed to allow for later analysis.

2.2.3 Equipment:

Owners were asked to complete PANAS (Sheppard and Mills 2002) and DIAS (Wright et al. 2011) questionnaires for their dog (as described in Chapter 1) prior to the start of the study.

Heart Rate Monitors (Polar RSX800c) were used to collect heart rate data from the dogs during their searches. See figures 2.7 & 2.8 below for attachment methods.

Figure 2.7. Illustration of fitting of heart rate monitor, held in place using vet wrap. The monitor was attached to the dog using the chest band around the dogs chest, with the monitor placed just under the left foreleg (using ultrasound gel to establish a connection), and a band of vet wrap tied from the left to the right side of the monitor across the front of the dogs body (seen in blue in the photograph). The Heart Rate monitor receiver was attached to the back of the chest belt between the dog’s shoulder blades to ensure connection was maintained between the monitor and receiver at all times.

Figure 2.8. Illustration of the heart rate monitor receiver fitted onto heart rate monitor belt to enable reading while moving without breaking connection.

Video Cameras (Panasonic ZXRs) were used to record conditions and search trials.

2.2.4 Pre-Assessment:

Owners were split into two groups of twelve to attend a pre-assessment with their dogs. They were informed that their dog would be needed for a full day of assessments which would be with 11 other dogs, and if they passed the assessment session, they were invited to attend a half day testing session with their dog 1-4 weeks after the pre-assessment.

All dogs were already trained by their owners to detect catnip on a variety of small scented material based articles. Each owner was asked to bring a selection of 10 pre-scented articles of similar size for their dog, in a sealed container that had all been treated with the same batch of the catnip scent at the same time, at least 1 week prior to the training day.

During the day, the dogs were kept in a group in a room 23ft x 23ft with their owner which will be referred to form herein as the ‘break room’, in their own area, with their bed and any chews or toys the owner brought for them to keep them comfortable and occupied while waiting for their pre-assessments. This room was next door to the search room where the searching took place but both rooms were identical in size.

When not practicing searching, in the break room the dogs were habituated to wearing a heart rate monitor by placing the monitor onto the dogs (as shown in Figures 2.7 & 2.8), then allowing the dogs to go for a walk and engage in play with their owner outside the building for 10 minutes, then allowed to rest for 10 minutes back inside the break room, following this their resting heart rate was recorded while the dog was sitting down. The first reading was taken after 5 minutes of wearing the device following the rest period, 3 recordings in total were taken at 1 minute intervals, and the average of these was recorded as resting heart rate for each dog.

In the search room, 20 card board boxes (sizes varied, no larger than 1ftx1ftx2ft, that previously did not contain food) were laid out for the dogs to search, boxes were placed within the search area which was divided into 9 sectors measuring 5ftx5ft, with discreet markers (numbering) on the floor as to where the boxes were placed (see figure 2.9 below).

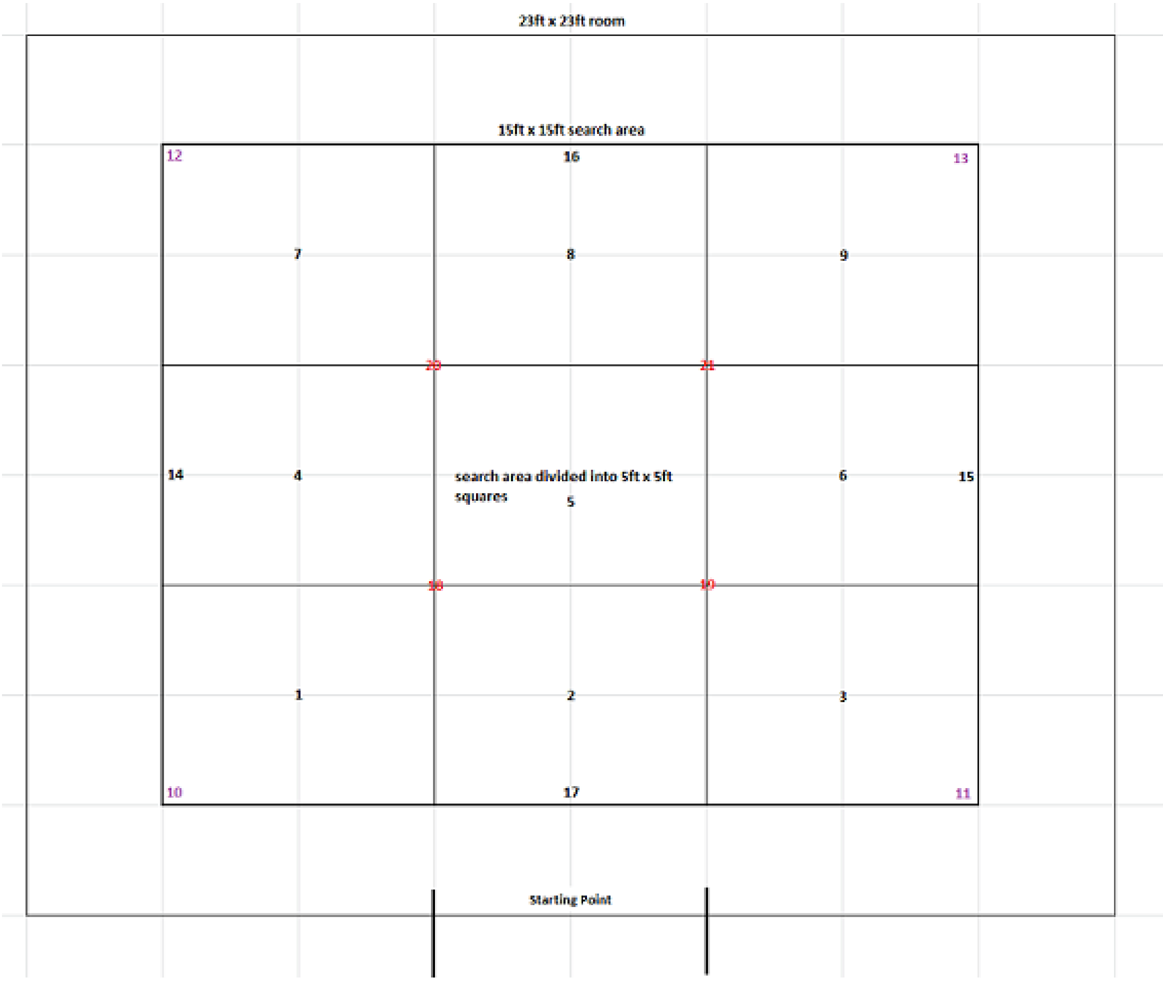


Figure 2.9 Diagram of search area. The room was divided into 9 quadrants, and numbers were written onto the floor in chalk to allow the researcher to know where each box needed to be placed

for each search. 20 boxes were used in each search, so one numbered position was always left blank (without a box).

Smaller boxes measuring 20cm x 20cm were placed within every larger cardboard box. These containers were used to conceal the scented articles and were changed for each search to avoid scent contamination in the following searches. While a smaller box was placed into every larger box, only certain smaller boxes contained a scented article. For a full plan of which boxes contained scented articles within each of the searches see Appendix 1a.

Each dog was given the opportunity to perform 3 searches on the pre assessment day, dogs were worked in order and then repeated so each had approximately a one hour break between their searches. Owners were given advice and guidance to ensure they were consistent with the pattern they followed when guiding the dog's search in the form of a handout (See Appendix 1b).

The position of the boxes remained the same for each search but the boxes which contained articles were discarded and changed for a similar box between searches. In each search there was 1 elevated hide (a hide is defined as a box that contained the smaller box in which an article was concealed) this was on top of a chair, 1 upside down hide (with the 20cm cardboard box containing the article placed with the opening against the floor), and one upright hide (with the 20cm cardboard box containing the article inside an upright box). There was also always 1 search quadrant that was left blank (with no box in at all). The location of articles were pseudo randomised across the 3 searches each individual dog participated in, but all dogs were exposed to the same 3 searches as each other. The boxes which contained scented articles were changed after every search. The entire search area was contaminated by distributing the scent on the article around the entire floor of the search area, and on all of the boxes.

Owners were informed that when their dog had indicated on a find, they were to reward their dog and help it access the find if appropriate, then replace the box that had contained the find (leaving it empty), pass the article to the tester (where it was sealed in a zip lock bag), then carry on with their search. For the training searches there were always 5 articles in the first search, 8 articles in the second search, and 5 articles in the third search for each dog. Owners were blind to the location and number of articles within the search. Owners were instructed to inform the tester when they thought they had cleared the test area, or timed out after 10 minutes of searching.

Dogs which allowed the heart rate monitor to be fitted, engaged in toy play, allowed a food bowl to be taken away from them, and found a minimum of 2 articles in each search on the training day were kept on for testing. For the training day owners brought in their own scented articles, on the training day owners were given a bag of 60 articles (per dog) of the same size and shape (cuts of towel measuring 5cm long by 3cm wide) to treat with their own scent at home and bring in for the testing day.

2.2.5 Testing:

Each dog was brought into the same building as the pre-assessment day on an individual basis and was fitted with the heart rate monitor in the break room and given 5 minutes to adjust to this (and for the researcher to obtain a baseline heart rate, checked against the reading obtained on the pre-assessment day).

Heart rate was measured throughout testing and the heart rate monitor was synchronised to the video camera prior to each test condition and each search by breaking contact between the monitor and the dog's skin, waiting for a blank reading on the monitor, re-establishing contact between the monitor and the dog's skin, then reading out the first 5 measures to the camera.

The dog was exposed to one of the pre-test conditions (see Figure 2.10 below) and then given the opportunity to search. The pre-test conditions were decided upon by a panel of experts in animal behaviour, and allocated arousal valence of either positive, negative or neutral, as well as placed onto a scale of what was thought to be the relative level of arousal the conditions would elicit compared to one another. They repeated this until they had been through all 10 conditions and completed all 10 searches.

Condition	***Valence (Arousal)***	***Level of Arousal***	***Proposed Arousal Scale (from 1 low to 10 high)***
Bed settle – the dog was asked to lie down on a bed for 5 minutes (based on Hiby *et al*, 2004, sending the dog to bed as a calm positive activity)	***+*** (The dog would be relaxing on their bed)	***Low*** (The dog was relatively still and not being interacted with)	***1*** ***(the dog will be relaxed)***
Neutral release – no warning or cues were given prior to the search, the dog was simply	***N*** (The dog was relatively still	***Low*** (The dog was relatively still	***2*** ***(The dog will not be doing anything but***

released into the search room. (Based on Call *et al* 2003, dogs are sensitive to the attentional focus of humans, a lack of attention and no engagement prior to the task, will put the dog in a neutral and only slightly aroused state).	and not being interacted with)	and not being interacted with)	***likely awaiting instruction)***
Attention Switch- Owner and researcher engaged in 2 minutes conversation while the dog stood on lead next to them, then the dog was suddenly released and cued to search (based on Call *et al* 2003, dogs are sensitive to the attentional focus of humans, a sudden switch should increase their arousal) .	***N*** (The dog was relatively still and not being interacted with)	***Low*** (The dog was relatively still and not being interacted with until the last moment)	***3*** ***(The dog will not be doing anything but will then get interaction in form of cue to search)***
Obedience- the owner carried out 2 minutes of obedience training with their dog, repeatedly asking for sit and down position, then sent the dog straight in to search (Based on Knol *et al* 1988, basic obedience tests used to assess guide dogs designed to create an alert yet neutral state in the dog.)	***N*** ***(The dog was not being rewarded or punished during the training)***	***Low/Moderat e*** ***(The dog was moving and responding to cues)***	***4*** ***(The dog was having interaction with the owner during the condition)***
Steady walk around room (steady walk)- the dog was	***N***	***Moderate***	***5***

walked on lead around a second room that they were not searching in for 2 minutes then sent in to search. (Based on Knol *et al* 1988, basic obedience tests used to assess guide dogs designed to create an alert yet neutral state in the dog).	***(The dog was not being rewarded or punished during the circuit)***	***(The dog was moving constantly throughout the condition)***	***(The dog was moving constantly throughout the condition)***
Agility jumps (jumps) - the dog was asked to jump 5 small agility jumps in a row, then sent straight into the search area and cued to search. (Based on Pastore *et al*, 2011, agility is generally enjoyed by dogs, but does cause increased arousal).	***+*** ***(The dog was interacting with the owner and was positively encouraged over the jumps and praised verbally)***	***Moderate*** ***(dog moving quickly and having a positive interaction with the owner)***	***6*** ***(more interaction with the owner, faster paced activity)***
Take away food bowl while eating (food bowl)- Owners brought a sample of their dogs own food, allowed the dog to start eating, then remove the bowl half way through and took the dog to search. (Based on van der Borg *et al*, 1991, shelter dogs assessed by taking away food bowl while eating to see how the dog responds to a moderate to highly arousing negative event).	***-*** ***(The dog had a positive experience (eating) interrupted and removed, generating frustration or negative emotional state)***	***Moderate/High*** ***(The switch from a positive to a negative state would increase arousal)***	***7*** ***(the emotion linked to food and eating then having this taken away would create a higher arousal)***

Are you ready? (gear up)- Owners verbally encouraged their dog for 1 minute prior to being sent in to search. (Based on Kubinyi *et al,* 2003, where owners used verbal encouragement to get their dogs motivated).	***+*** ***(Owner using excited voice, praise and verbal encouragement to get the dog excited)***	***High*** ***(high intensity interaction with the owner)***	***8*** ***(high intensity interaction with the owner)***
Tease with toy- The dog was teased with a toy, not being able to take hold of it for 1 minute, the toy was then thrown out of reach and then the dog sent into the room to search (Based on Hiby *et al*, 2004, using play as a highly arousing positive activity in shelter dog assessments, followed by withholding the toy to induce frustration to assess the dogs response to a highly arousing negative emotion).	***-*** ***(The dog will be frustrated by not being able to interact with the toy)***	***High (motivation for the toy is high-as testing in pre assessment, so being teased with it would create a high level of frustration)***	***9*** ***(as motivation for the toy is high, and being prevented from obtaining it)***
Play with toy-maximum excitement possible (play max)- Owners were informed to play with their dogs to get maximum excitement, allowing their dogs to take hold of the toy and rag on it for 1 minute then the toy was swapped out for a treat and the dog sent in to search. (Based on Hiby *et al*,	***+*** ***(The dogs motivation for the toy is high, the owner is positively encouraging the dog and praising the dog for play,***	***High*** ***(Dogs being allowed to get very excited and having lots of positive rewards)***	***10*** ***(Dogs being allowed to get very excited and having lots of positive rewards)***

2004, using play as a highly arousing positive activity in shelter dog assessments).	***and the dog is being allowed to play, the toy is then swapped out with food, another positive for the dog)***		

Figure 2.10. Pre-testing conditions for each search, the valance of the arousal and the supposed level of arousal each condition was thought to elicit (‘+’ indicates ‘positive’, ‘-‘ indicates ‘negative’, and ‘N’ indicates ‘neutral’). Rationale behind the allocation of the valence, level and order of arousal is also included within the table.

The conditions occurred out of sight of the search in the break room, breaks were given after each search and before the next condition to allow heart rate to return to around its resting value. Once this had occurred the dog was given a minimum of 2 minutes in this resting state. The first 12 dogs were each allocated a different order of conditions, the second group of 12 dogs were paired with a dog from the first group that most closely matched their level of training and experience, and given the conditions in the reversed order to the first dog (see Appendix 1c. for order of conditions), in order to provide a degree of counterbalancing control.

The owner was informed that in each search there would be between 2 and 8 items for their dog to detect, and that it was up to them to inform the researcher when they felt the dog had cleared the search area. However, unknown to the owner there were always 5 items for the dog to detect in each search to enable comparison between search performances under different conditions. They were also informed to allow their dog to free search until the point where they felt the dog would no longer make progress in the search unaided, when they thought this was the case they were to switch to a directed search (see appendix 1b for the instruction given for directed search pattern).

The location of articles were pseudo randomised across the 10 searches each individual dog participated in, but all dogs were exposed to the same 10 searches as each other, in the same order (See appendix 1c). The boxes which contained scented articles were changed after every search. The entire search area was contaminated by distributing the scent on the article around the entire floor of the search area, and on all of the boxes.

Owners were informed that during a search, if their dog identified a false positive, they were to replace the box where it was found. This was to allow the dog to re investigate if they chose to.

Between tests the cardboard boxes were switched around and when laying articles every box was touched to ensure the human scent was distributed around the search area. The cardboard boxes which had contained articles were all changed following every search. The location of the hidden articles was pseudo-randomly changed for each test search from search 1 to search 10 (see appendices 1a and 1c), as was the nature of the hides, in that 2 search locations were always in an elevated position (on top of a chair). Each dog was given the same formation of boxes, locations of hides, blanks and raised platforms as all the other dogs (see appendices 1c). This ensured that on all of the searches every dog saw the same pattern and had the same location of hides for search 1, this then changed for search 2, but again all dogs had the same pattern, and so on for all the following searches.

As part of the consent form owners agreed to not discuss the testing scenarios or their dogs' performance until all dogs had been tested, at which point their result would be released to them and they were free to discuss the searches.

2.2.6 Measures:

Performance was defined by breaking the search into 2 components (free search and directed search) and measured by weighting scores so that more value is given to articles that were located under more difficult circumstances (for example if the dog has already located 4 articles in the free search, there is only one article left to locate so there is less chance of finding the remaining article during the directed search), using the formula:

a+(b/2(5-a))

Where (a) is the free search and (b) is the directed search.

Optimal performance for each dog was defined by the search in which they received the highest score.

If the Yerkes Dodson prediction dominates performance it is anticipated that pre-test condition should be a significant predictor of optimal performance. If the IZOF dominates, then arousal (measured by HRV) during the free search and/or precondition should significantly predict optimal performance of each dog, regardless of pre-test condition.

2.2.6.1 Data Analysis

Data were initially visually inspected. Graphical representations of arousal or pre-test condition versus performance were generated to look for trends reflecting either IZOF or Yerkes Dodson Law effects.

Data were analysed using statistical package SPSS 21.0.

It was anticipated that HRV data would be analysed to tests the IZOF hypothesis, unfortunately, due to problems with heart rate recording, an alternative analysis was undertaken to further test the Yerkes Dodson hypothesis instead. The data set was divided into two, (high versus low arousal conditions, with a clear separation between the two (i.e. moderate conditions excluded) and it was determined if dogs performed best in one of these categories as a whole using matched pairs t-tests based on their average performance across the three tests in each category. Pearson's correlations were used to assess the strength of relationships between pre-test conditions (based on putative arousal) and performance (defined by the formula above), as well as correlations between performance across conditions.

To attempt to test for the effect of individual differences on search performance, differences between subjects average search scores were examined in terms of the dogs scores for elements within their PANAS (Positive Activation, Energy & Interest, Excitement, Persistence, Negative Activation) and DIAS (Overall Questionnaire Score, Responsiveness, Aggression & Response to Novelty, Behavioural regulation) profiles, using Pearson correlations to cautiously examine the impact of individual differences on performance.

2.3 Results

All of the dogs that went through the training day went through to testing. 3 dogs were excluded from the study because they failed to complete all 10 searches during testing, they were all withdrawn by their owners due to fatigue. HRV parameters could not be used due to insufficient data collection resulting from errors in the monitors reading the heart rate during movement. Over 50% of the data contained over 5% of errors, too many to be considered reliable (Schoberl, *et al* 2015). For two of the dogs data was irretrievable due to HRM malfunction. See Figure 2.11 below.

Search	1	2	3	4	5	6	7	8	9	10
Dog	HRV (pNN50)									
1										
2										
3	25	17	15	17	21	19	23	33		31
4	27	28	37		26	31	49			12
5	12	7	7	5	9	22	12	13	9	10
6	4	6	7	10		18	42	33	36	43
7	31									
8		41			16	18	18	18	22	21
9						0.1	0.2	1.2	0	0.7
10							4	8	12	7
11	2							18	32	14
12	37	42	46	32	49	51	43			
13	22	12	15	13	17		24	25	35	38
14	33		32							
15	4	9	8	13						
16	22	23	23	28						
17	19	11	38	32	17	16		16	13	9
18	5	10	11	15		49				
19	18	19	20	24	19	20	15	21	25	31
20			24			38				

Figure 2.11. Extent of data loss from HRM for all of the dogs in all of the conditions. Red indicates error rate over 5%, orange indicates error rate up to 5%. Grey indicates data irretrievable. HRV value is pNN50 (percentage of successive beats that differ more than 50 ms from each other), so the higher this number the more variability. All values are rounded to whole numbers. Key: HRV= Heart Rate Variability.

2.3.1 Conditions and Individual Performance

Individual dog performance varied across and within the conditions for each dog (see Figure 2.12 below).

Condition/ Arousal		Average Search Score	Performance score for each individual dog																
low	1. settle	2.28 +/- 1.3	1.33	4.06	4.06	2.14	3.13	0.50	0.50	2.29	2.21	3.13	2.21	5.00	2.14	2.21	1.33	0.40	2.14
	2. neutral	1.93 +/- 0.9	3.13	2.21	2.21	3.13	1.08	1.33	1.33	1.33	1.33	3.13	2.14	3.13	1.17	3.13	1.33	1.33	0.40
	3. switch	2.42 +/- 1.2	1.33	3.13	4.06	4.06	2.21	1.17	2.14	1.25	1.33	1.33	4.06	3.13	2.14	3.13	1.33	1.33	4.06
	4. obedience	2.3 +/- 1.3	2.21	4.00	4.00	4.06	3.13	2.14	1.25	1.33	1.25	3.13	3.13	3.13	0.50	2.21	3.25	0.50	0.50
medium	5. walk	2.65 +/- 1.4	3.13	4.06	4.06	4.06	1.25	1.33	1.25	1.25	0.50	3.13	3.13	5.00	3.13	4.06	1.33	1.25	3.13
	6. jumps	2.64 +/- 1.1	3.13	4.06	4.00	4.06	3.06	1.33	1.25	1.33	1.25	2.21	1.33	3.06	2.14	2.21	3.13	4.22	3.13
	7. food bowl	2.59 +/- 1.4	3.13	3.13	2.00	4.06	1.25	1.33	0.40	2.21	1.33	3.13	3.13	5.00	2.21	5.00	2.21	0.50	4.06
High	8. ready	2.1 +/- 1.1	1.25	4.06	3.13	2.21	2.21	1.33	1.25	1.33	0.50	3.13	4.00	2.14	0.40	3.13	2.21	1.25	2.21
	9. tease	2.26 +/- 1.2	3.13	4.06	3.06	2.21	3.13	2.21	0.40	1.33	1.33	4.06	4.06	2.14	1.17	2.21	1.33	0.50	2.14
	10. play	2.58 +/- 1.3	1.25	5.00	4.00	2.21	3.06	1.33	1.33	4.00	3.06	3.13	3.06	4.06	2.14	3.13	1.33	0.50	1.25

Figure 2.12. Individual weighted scores for each dog in each of the searches following the separate conditions. Green indicates the highest scoring searches for each dog, and red indicates the lowest scoring searches for each dog to allow for visual inspection of any trends in individual dog performance. Weighted scores were calculated using the formula 'a+(b/2(5-a))', where (a) is the free search and (b) is the directed search.

A graphical representation of pre-test condition (ordered from left to right in terms of increasing inherent arousal level) and performance on a group and individual basis were generated to look for trends resflecting either IZOF or Yerkes Dodson Law effects, see figure 2.13 below. It appears that search score increases across the population as the levl of arousal increases, but not significantly so.

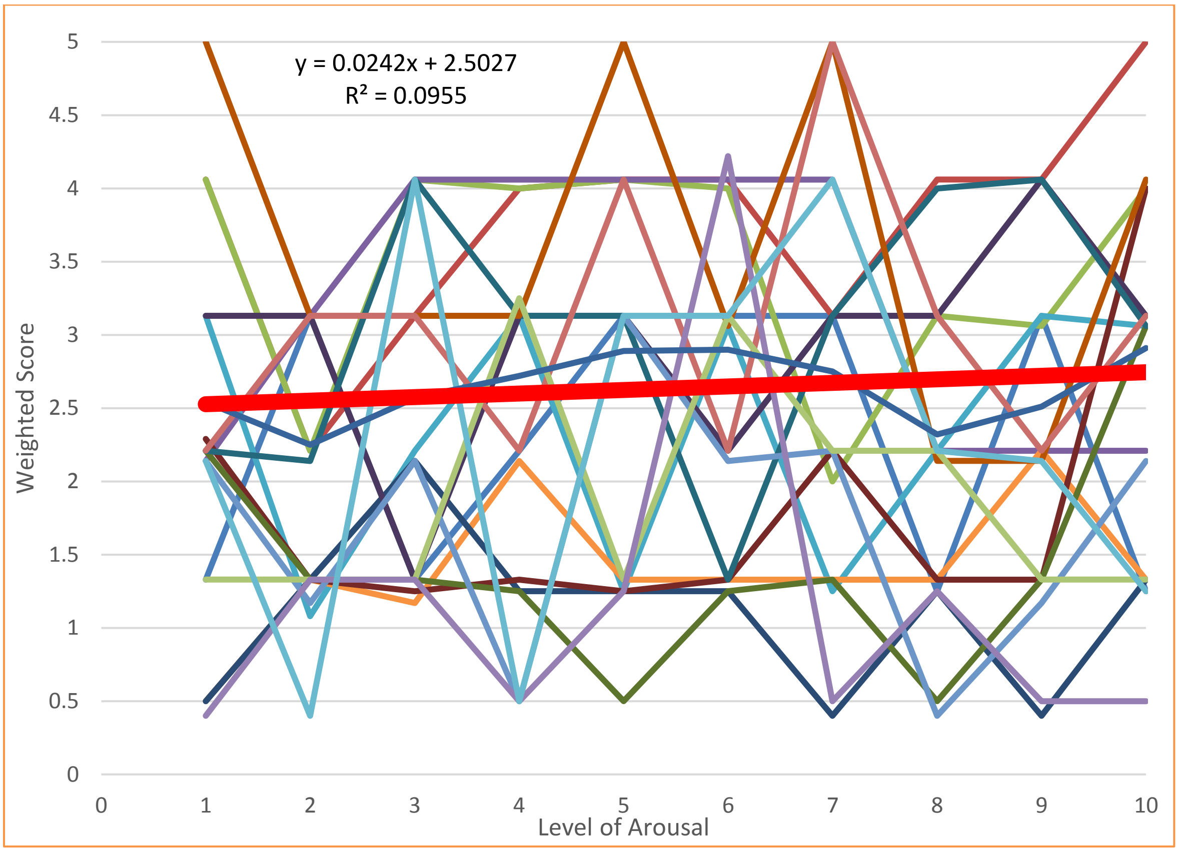


Figure 2.13. Individual dog performance (weighted score) in each of the conditions ranging from 'bed settle' (low level putative arousal (1)) through to 'play with toy-max excitement' (high level putative arousal (10)). Weighted scores were calculated using the formula 'a+(b/2(5-a))', where (a) is the free search and (b) is the directed search. The red line is the trend line showing a slight increase in performance as arousal increases.

2.3.2 Arousal and Performance

The conditions were split into two groups of 3, the lowest arousal 3 and the highest arousal 3 to be used for analysis of high and low arousal against average performance to determine if there was an optimal level for performance across the population associated with either high or low arousal. Data met normal assumptions unless otherwise stated.

Matched pairs t tests found no significant differences between performance in the low arousal condition set and either the high arousal (t=-0.666, p=0.515) condition set or the average overall search score (t=-684, p=0.504). No significant differences were observed between the high arousal condition set and the average search performance score (t=0.647, p=0.527). Performance in the high and low condition sets was found to be significantly correlated (Pearsons Correlation 0.776, p=<0.001), as were the low condition set and the average score (Pearsons Correlation 0.932, p=<0.001), and the high condition set and the average score (Pearsons Correlation 0.952, p=<0.001). See Figure 2.14.

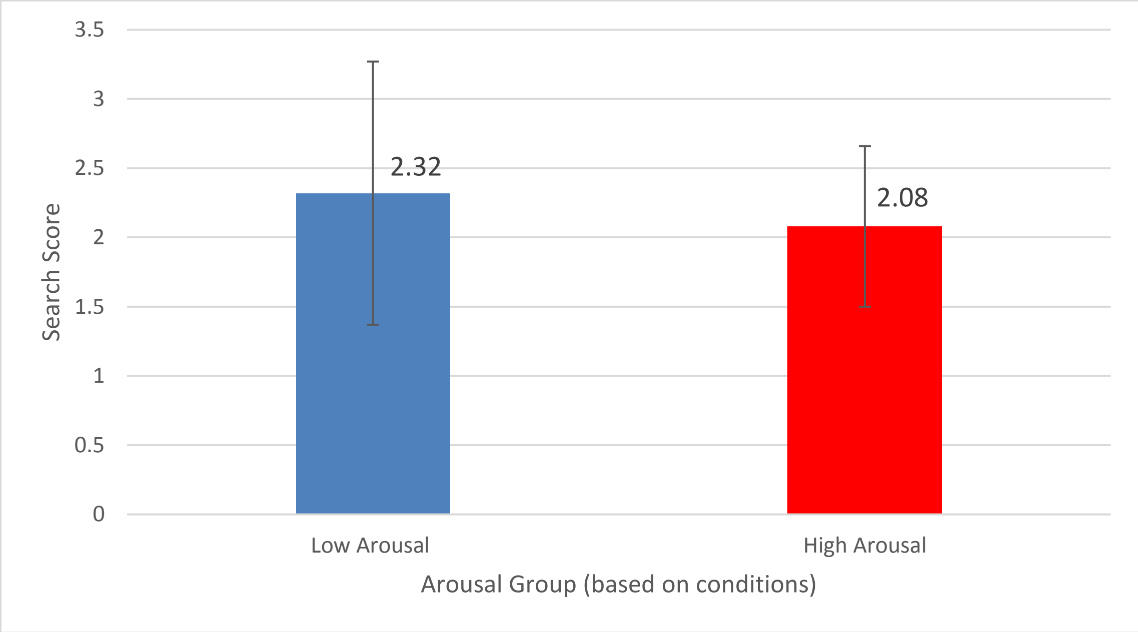


Figure 2.14 Average (mean) performance for all dogs in the conditions grouped into high and low arousal eliciting conditions, including standard error.

2.3.3 Temperament and Performance

Individual differences in terms of scores from PANAS and DIAS questionnaires were analysed against average performance score to determine if temperament factors correlated with performance using Pearson's Correlations.

Correlations were found between average search score and PANAS component Negative Activation (0.609, p=0.009), and average search score and OQS showed a borderline significant correlation (0.409, p=0.051). See Figure 2.15.

Element from PANAS or DIAS	Pearsons Correlation (p) between average search score and element from PANAS or DIAS
PANAS: Negative Activation	.609** (0.009)
PANAS: Positive Activation	0.356 (0.161)
PANAS: Energy & Interest	0.2822 (0.272)
PANAS: Excitement	0.213 (0.411)
PANAS: Persistence	0.229 (0.378)
DIAS: Overall Questionnaire Score (OQS)	0.48 (0.051)
DIAS: Responsiveness	0.409 (0.103)
DIAS: Aggression & Response to Novelty	0.389 (0.122)
DIAS: Behavioural Regulation	0.114 (0.664)

Figure 2.15 Analysis of correlations between average search score across all ten conditions and temperament traits from the PANAS and DIAS.

For both correlations the trend was that an increase in the score of the trait measured by the PANAS or the DIAS correlated with an increase in average search score. See Figure 2.16 below for correlation with negative activation and Figure 2.17 below for correlation with OQS.

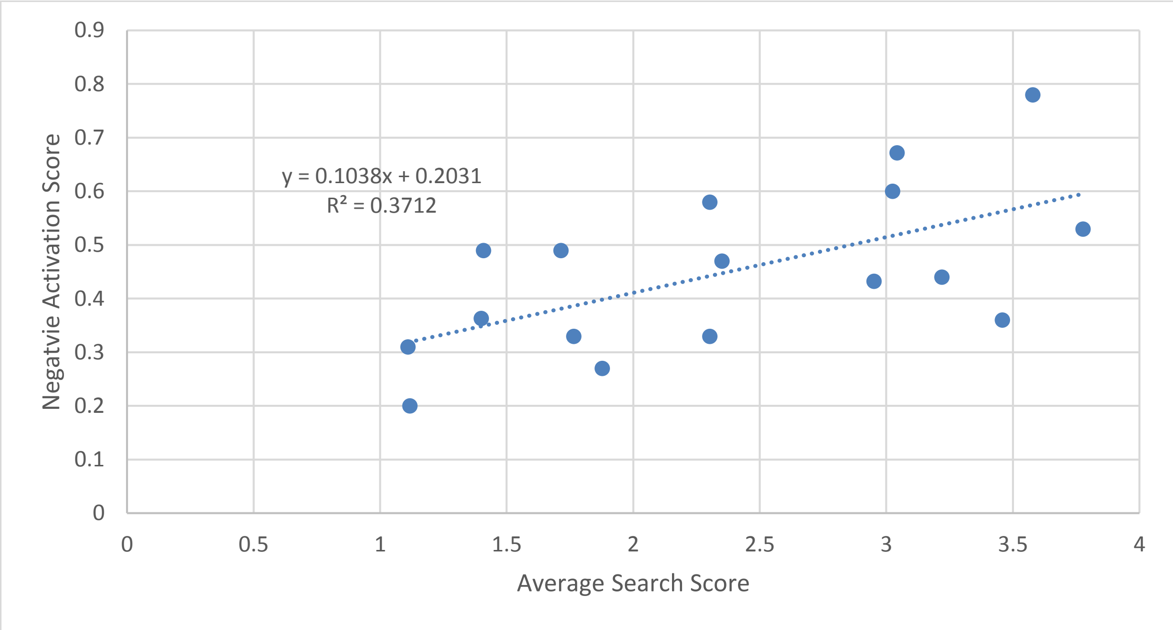


Figure 2.16 Negative Activation Score (PANAS) compared to Average search score across all conditions for each dog. Where average search score has been taken as the mean score from all 10 searches.

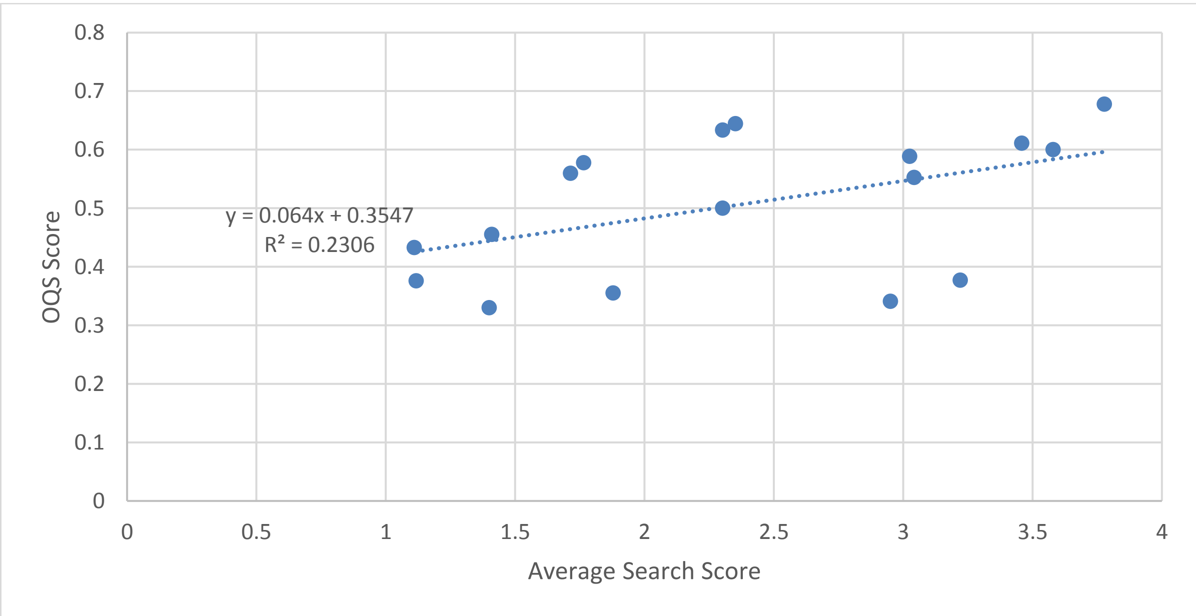


Figure 2.17 Overall Questionnaire Score (DIAS) compared to Average search score across all conditions for each dog. Where average search score has been taken as the mean score from all 10 searches.

2.4 Discussion

Dogs appeared to score differently to each other in terms of their performance regardless of the condition they were exposed to. Currently, methodologies to quickly assess physiological measures of arousal are limited and require further refinement to be useful in terms of monitoring field performance. The method in this study may have failed as the HRM used has only previously been validated for assessing HR and HRV in stationary dogs (Jonckheer-Sheehy *et al*, 2012). As HRV measures could not be used there are risks associated with inferring the level of arousal each condition elicits in a dog as all dogs will experience arousal differently. It is particularly important to consider past experiences. It has been shown that dogs trained using punishment methods have a higher incidence of behaviour problems compared to those trained using reward based methods, suggesting that training methods have an impact on future behaviour, and punishment can lead to anxiety based problems (Hiby *et al*, 2004). Some dogs used in this study for this thesis may have a negative association with being sent to bed (one of the conditions being a bed settle, so this may create negative high arousal in the dog if it is associated with punishment), and obedience may have been trained using lots of food rewards, which could create positive valence high arousal, or using punishment methods, which could create a negative valence. This is a flaw in the methodology of this study, which would have been countered by the use of HRV data to establish real physiological arousal, had the data been usable. As it wasn't attempts were made to make the conditions useful in terms of giving an idea of the level of arousal, but it is acknowledged that there are issues with this method that may have resulted in a lack of usable results in terms of the impact of arousal on performance. As this was the case measures for individual differences were turned to in an attempt to explain the differences in performance.

If task arousal had an optimum level, and the conditions elicited the arousal they were purported to, a distinct curve might have been expected to have been observed in terms of search performance peaking for all dogs within a certain range and dropping off either side of this, or at least some consensus in terms of a precondition or group of preconditions (high verses low arousal) which favoured performance. This was not the case so the relationship between arousal and performance is likely more complex than this, and potentially linked to the individual. There are also a number of methodological reasons as to why arousal may not correlate with performance, potentially the conditions did not elicit the level of arousal anticipated, or the levels of arousal elicited from the conditions was not spread far enough apart in terms of intensity, which was prevented due to ethical considerations.

Negative Activation as measured by the PANAS was correlated with search performance and interestingly the higher the score for negative activation the higher the performance score suggesting that individuals that are more sensitive to potentially threatening or aversive qualities within the environment overall performed better in the search tasks compared to those with lower scores. This could be because those individuals predisposed to notice the negative stimuli are also at a higher baseline of arousal and more attentive to change in the environment, which is an integral part of the search task, even if this has been positively reinforced in training. It may be that temperament is an important mediator of performance, or acts as a buffer for arousal as has been shown previously in the case of impulsivity in terms of the impact of differing levels of arousal on task performance in more or less impulsive humans (Anderson, 1994). With impulsive individuals under low arousal conditions showing a deterioration in performance in a difficult task under a high arousal condition, while non impulsive individuals performance under a high arousal condition first improved an then deteriorated at the same task (Anderson, 1994), this highlights the role of temperament traits in moderating behaviour. Research by Bray *et al*, 2015 compared pet dogs (considered the high arousal group) to guide dogs (which are trained for low arousal) in a detour task which involves inhibition. As a cognitive task lower levels of arousal would result in better performance according to Yerkes-Dodson law. When the dogs were encouraged by a human (attempting to increase arousal levels) dogs in the low arousal group (the working dogs) performed better in the task (negotiating the detour faster) compared to dogs in the high arousal group (pet dogs). This suggests that temperament plays a role in moderating arousal and performance (Bray *et al*, 2015). However, temperament was not assessed as an underlying factor in this study and as guide dogs are bred and selected for certain traits this may have had an impact on their response or ability to complete the task.

In the present study it appears that each dog is performing differently to one another, despite being subjected to the same conditions, this would indicate that the preconditions may be experienced differently by different dogs or that arousal as elicited by the pre-conditions did not affect task performance. As individuals performing the same task in the present study, even under the same conditions performed differently, this could suggest that the individual and how they experience arousal as well as how this links to performance is potentially more important than task arousal. In the human literature self-report is frequently used to determine anxiety levels prior to performance and ensure that the conditions they were exposed to elicited the desired response (Ruiz *et al,* 2015. Ruiz, Hanin & Robazza, 2015). In non-human animals the difficulty in measuring and assessing arousal levels lies in the methodology. It is also important to consider the limitations of not having a physiological measure of arousal in the study, so the arousal levels thought to be generated in the conditions were putative, and potentially not representing the actual arousal generated. This needs to be further addressed to be able to move forward in this field. As differences were identified in some of the temperament components of core affect and impulsivity between the subjects and linked to performance, this could support the idea that optimal performance is more related to temperament at either a task or arousal level. There is also the potential for task performance to be a result of task and individual arousal, as IZOF and Yerkes-Dodson laws are not mutually exclusive. It could be that the population recruited as trained search dogs naturally fell into the range of task arousal best suited for search work, or had an optimal temperament for searching in this way as their owners probably would not have persisted with the training of their dogs to this level if they were not succeeding. Their success could be a result of training, or natural aptitude.

Developing an understanding of factors that affect performance and establishing ways to predict or manipulate arousal to impact upon performance could help with ensuring dogs are in an optimal arousal state before going into work to ensure performance is consistently high in the field. There is a need to consider each dog as an individual in terms of performance and find out what works best for each dog. Repeat testing of dogs would help establish if the pre-task conditioning effects were consistent within individuals and the relationships between different preconditions reliable.

Because of the limited results from this study as a result of problems with being able to physiologically measure arousal, and an apparent lack of relationship between performance and pre-condition, this line of enquiry was not pursued further, leaving it unknown if the results would have been shown to have reliability. A lack of physiological measure for arousal also meant that it was not possible to determine validity of the conditions and the suggested levels of arousal they elicited. However, this study does highlight the importance of recognising individuality and the sources of individual variation. Therefore in the next Chapter a review of current selection processes is undertaken to focus on identifying gaps in the literature when it comes to working dog selection processes to identify areas for further study.

Chapter 3: A Review of the reliability and predictive validity of temperament tests in assessing behavioural characteristics of relevance to working dogs.

Synopsis

Working dogs are selected based on predictions that they will be able to perform specific tasks in often challenging environments. Selection processes aim to minimise the loss of working dogs from active service due to an inability to perform their job. However, withdrawal from service in working dogs is still a problem, bringing into question the effectiveness of the selection processes. A review was undertaken aimed at bringing together available information on the reliability and predictive validity of the assessment of behavioural characteristics relevant to working dogs to establish gaps in the literature where tools could be developed for use to aid prediction of success in working dogs. The search procedures resulted in 15 papers being included for evaluation. Given that a large range of behaviour tests and parameters were used in the identified papers, firstly behaviour tests were grouped by their associated underlying traits based on positive core affect (willingness to work, human-directed social behaviour, object-directed play tendencies) and negative core affect (human-directed aggression, withdrawal tendencies, sensitivity to aversives). Secondly, papers were examined for reports of inter-rater reliability, within-session intra-rater reliability, test-retest reliability and predictive validity. The review revealed a variation in testing methods and inconsistencies in terminologies, study parameters and indices of success. There is a need to standardise the reporting of these aspects of behavioural tests in order improve the knowledge base of what characteristics are predictive of optimal performance in working dog roles, improving selection processes and reducing working dog redundancy. The review highlighted a need to identify additional measures to use for assessing working dogs to reduce the rate of withdrawal and failure to certify.

3.1 Introduction:

3.1.1 Selection Tests

Selection tests are frequently used to determine individuals likely to succeed in certain roles or environments, equally they aim to eliminate unsuitable individuals. In terms of working dogs there is a need to select individuals with not only the ability to perform specific tasks, but also those who are robust enough to cope with a working role in an often challenging environment.

It is important to note that there is a difference between assessing behaviour and temperament. Behavioural tests can be developed and used to predict behaviour but may be contextually limited, while temperament tests are based on inferred traits derived from measurements of behaviour across a number of contexts (Taylor & Mills, 2006).

3.1.2 Assessing temperament

As discussed in Chapter 1, it is important to distinguish temperament from personality or character, as the terms are sometimes used interchangeably in practice, although there are slight differences (Horwitz & Mills, 2009. Jones & Gosling, 2005).

The assessment of temperament in dogs is of interest not only to research scientists in terms of developing an understanding of their behaviour (Gosling, 2001), but also to professional animal handlers, breeders and trainers (Jones & Gosling, 2005). As evident from other species, depending on whether the purpose of assessment is for welfare evaluation (Dawkins, 2004. Pritchard *et al*, 2005), predicting behaviour (e.g. in companion animals, Poulsen *et al*, 2010), understanding risk (Haverbeke *et al*, 2009) or to select for specific qualities (Serpell & Hsu, 2001), the form of the assessment may vary (Gartner, 2015). For example some assessments may take place over a number of days, others within the space of a single session. Temperament, character and personality have been extensively studied in dogs (Gartner, 2015). Assessments usually consist of examining the way an individual responds in a set of situations designed to resemble those it is likely to encounter throughout life, often with the aim of predicting future behaviour in similar circumstances (Svartberg, 2005). By assessing behavioural responses across a range of contexts, a behavioural profile is developed of the dog which may be interpreted in terms of character, personality or temperament (Barnard *et al*, 2012).

Assessments for different purposes may also focus on different behavioural tendencies. For example, shelter dog assessments tend to focus on ensuring the dog will be safe and suitable in a home environment (Bollen & Horowitz, 2008), aiming to match dogs to appropriate homes (Jones & Gosling, 2005). While working dog assessments may focus more on traits like confidence and "play drive" (Svartberg, 2002). Working dogs need to be able to cope in a variety of situations, which have the potential to be stressful. Because of this, potential working dogs should be assessed and selected based on a combination of both ability and temperament. Ability relates to the dog being able to do the work required of it and this quality may largely focus on health, learning and the functioning of the sense organs enabling the dog to complete tasks and so does not always predict performance in challenging environments. Temperament consists of differences in the behaviour of individuals which are relatively consistent over time and across contexts, that are grounded in emotional states or core affect and other behavioural regulatory processes. It is a measure of individual sensitivity to certain qualities in the environment and so is a potentially important predictor of performance in complex environments (Slabbert & Odendaal, 1999).

Behaviour tests, such as those mentioned above may assess character (Trybocka, 2010), personality traits (Svartberg & Forkman, 2002) or temperament traits in dogs (e.g. De Meester *et al*, 2008 & 2011. Wilsson & Sundgren, 1998). In addition to behavioural tests, questionnaires may be used to assess individuality. Temperament in dogs can be assessed through questionnaires developed for psychometric profiling, such as the Positive and Negative Activation Scale (PANAS (Sheppard & Mills, 2002), or the Dog Impulsivity Assessment Scale ((DIAS) Wright *et al*, 2011). Instruments such as the CBARQ (Hsu & Serpell, 2003) generate a behavioural profile the Monash Dog Personality questionnaire is focused on defining canine personality. All of these are looking to identify individual differences, but refer to this using different terminology (e.g. temperament or personality). Questionnaires remove the need for implementing behaviour tests, but testing is useful when knowledge about the dog is not available to reliably complete questionnaires.

Scoring in behavioural tests can be objective, subjective, or a combination of these. In working dog selection, behavioural assessments are widely used to determine that genetic selection and early life experiences have resulted in individuals who appear both willing to work, and robust enough to cope with working in challenging or distracting environments (Haverbeke *et al*, 2009. Svartberg, 2002. Evans *et al*, 2007. Sinn *et al*, 2010. Slabbert & Odendaal, 1999). However, these assessments are not always predictive of long term performance in the field. This is evident from 82% of the reported dropout following certification, for military service dogs over the age of 5 being for behavioural reasons (Evans *et al*, 2007).

Temperament tests, regardless of the method used, should identify characteristics that are stable across time and context. In the case of working dogs, the trait should also be relevant to some aspect of performance and so be predictive of success (Wilsson & Sinn, 2012). In order to assess reliability and predictive validity of assessment tests for measuring behavioural characteristics relevant to working dogs, that are used to build temperament profiles, it is also necessary to understand which characteristics are important. In the working dog literature, as in the wider animal literature, it is clear researchers may use different terminology to describe similar behaviours e.g. Svartberg 2002 describes boldness, while DeMeester *et al* 2008 talks about confidence in a similar way. Most appear to use trait constructs when referring to temperament traits, but again the terminology differs depending on individual researchers and the fields in which they work (Gosling, 2001). Therefore the aim of the current review was to summarise the available research relating to the range of behaviour tests used for assessing behavioural characteristics relevant to working dogs by undertaking a review of the scientific literature and critically appraising and evaluating the data available in the published literature, irrespective of the terminology used.

3.2 Method:

3.2.1 Review Protocol

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were adhered to to perform this review (Moher, Liberati, & Altman, 2009). The inclusion criteria for selection of articles included (a) literature that reports behavioural observation tests to assess dog behavioural characteristics relevant to working dogs (b) articles written in English, (c) articles accessible via direct download or contact with the authors. Exclusions were not made based on dog characteristics (i.e. age, breed, and pet or working dog status) or test parameters. Reviews and meta-analyses were excluded. Questionnaire based studies were also excluded, since these did not satisfy the requirements of a 'behavioural observation test'. Studies which did not assess factors which may relate to working dog performance were excluded (e.g. papers purely focussing on heritability of traits), but it was not stipulated that the studies should include a working dog (or potential working dog) sample, since behaviour tests which have been applied to a non-working dog sample may be valuable for identifying future applications for use with working dogs.

3.2.2 Literature Search

Literature searches were conducted in electronic databases (PubMed, Scopus, Science Direct and Google Scholar) from their year of inception up to the end of November 2016. Figure 3.1 contains the list of search terms used. Search terms were decided following expert consultation with established researchers in the field and through evaluation of common terms used in titles and abstracts of papers known to the researchers.

Working dog temperament test(s)	Predictability of working dog behaviour
Working dog personality	Working dog performance
Working dog assessment(s)	Working dog selection test(s)
Service dog temperament test(s)	Service dog selection test(s)
Dog personality	Dog behavioural test(s)
Predictability of dog behaviour	Dog performance
Police dog temperament	Ideal working dog
Police dog behaviour	Police dog behaviour(s)
Police dog performance	Police dog selection
Military dog behaviour(s)	Military dog temperament
Military dog performance	Military dog selection
Assistance dog(s)	Assistance dog temperament
Assistance dog temperament test(s)	

Figure 3.1 Search terms used in the literature search for the review.

The search process was stopped when 10 consecutive electronic pages on the database being searched produced no relevant titles. At each stage of the review process, a selection of articles were cross checked by another researcher to ensure agreement on inclusion and exclusion decisions. Full text articles for all papers were sourced electronically, or through direct contact with authors. To assess the reliability and validity of the behavioural tests, further statistical information on the reliability of the reported behaviour tests was requested from corresponding authors of all the papers. For a full list of papers that were included in the review see Appendix 2. For a breakdown of the search process and where papers were rejected see Figure 3.2 below.

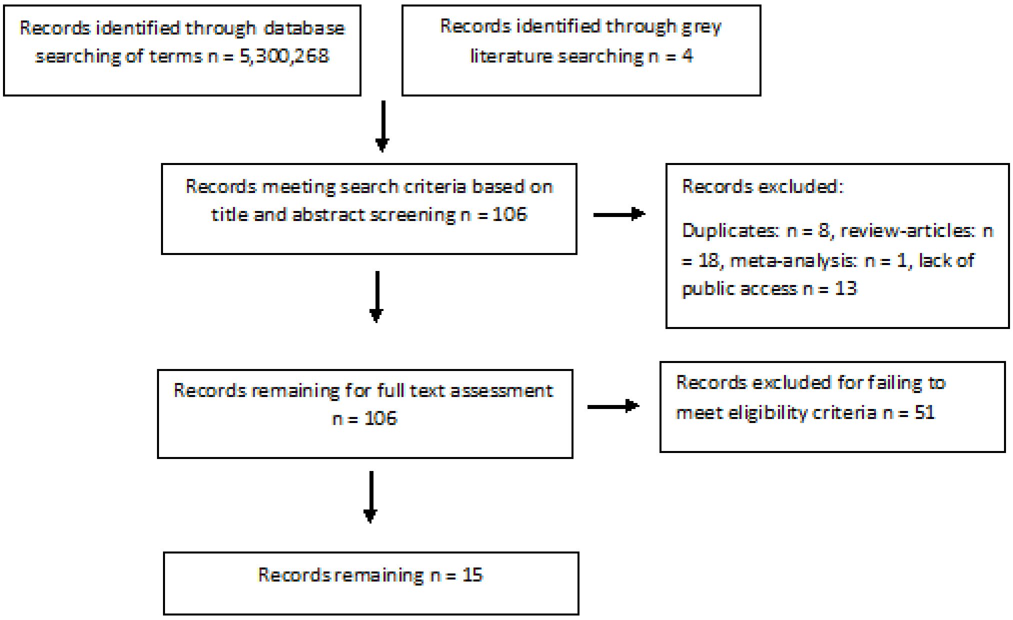


Figure 3.2 Flow chart illustrating the selection process of papers included in the review. Including the stages at which papers were rejected and the number of papers taken forward at each level of selection.

3.2.3 Defining Terminology

Papers were reviewed to identify the range of behavioural characteristics assessed in the selected literature and to categorise each behaviour into a thematic group, relating to underlying traits. Traits are typically inferred from behavioural tests and different tests were expected to label similar traits using different terminologies. To manage this, it was decided to classify the proposed traits assessed by the behavioural tests around a framework extending from their direct or indirect relationship with core affect (positive versus negative emotional states) (Taylor & mills, 2006), see Figure 3.3. Positive emotional states were characterised as those related to sensitivity to salient positive qualities in the environment, as observed through behaviours such as human, or object-directed play tendencies. Negative emotional states were characterised by tests relating to sensitivity to potentially aversive qualities in the environment, as observed through behaviours such as human-directed aggression and withdrawal tendencies.

Core Affect	Behavioural Characteristic	Examples of Terms Used	Examples of Test Parameter
Positive emotional state	Willingness to work	Search focus, motivation	• Searching for an object without interruption (Sinn et al. 2010; Slabbert & Odenaal, 1999). • Speed and hesitation with obstacle crossing (Asher et al. 2013; Slabbert & Odenaal, 1999; Svobodova et al. 2008). • Distraction behaviours shown when another dog passes by (Batt et al. 2008) • Ratings of willingness to return a ball (Wilsson & Sundgren, 1998).
	Human-directed social behaviour	Greeting behaviour, approach to strangers	• Ratings of willingness to greet a stranger (Svartberg, 2002; Svartberg et al. 2005; Svobodova et al. 2008; Wilsson & Sundgren, 1998). • Body posture / behaviour during approach, petting/examination (Asher et al. 2013; Svartberg et al. 2005)
			• Behaviours and vocalisations during toy play (Sinn et al. 2010) • Time to release toy (Sinn et al. 2010) and latency to catch toy (Batt et al. 2008).

	Object-directed play	Toy play, chase	• Intensity and interest in toy/tug-of-war (Svartberg, 2002; Svartberg, 2005; Svartberg et al. 2005; Svobodova et al. 2008) • Forceful toy play (Fuchs et al. 2005). • Immediate reaction to toy (investigate first or start to play) (Wilsson & Sundgren, 1998). • Responsiveness to toy versus assessor (Asher et al. 2013)
Negative emotional state	Human-directed aggression	Defence drive, stranger-directed aggression	• Posture, behaviour (Sinn et al. 2010; DeMeester et al. 2008, 2011) and vocalisations towards tester (Sinn et al. 2010). • Speed to bite and force of bite to tester (Sinn et al. 2010) • Level of aggressive response when provoked (Fuchs et al. 2005; Slabbert & Odenaal, 1999), startled (Svartberg, 2002; Svartberg, 2005; Svartberg et al. 2005), or approached (Sherman et al. 2015). • Time to taken to calm down after being provoked (Fuchs et al. 2005).
	Withdrawal Tendencies	Investigation-exploration	• Exploratory behaviour when startled, by visual or acoustic stimuli (Sherman et al.

			2015; Svartberg, 2002; Svartberg, 2005; Svartberg et al. 2005; Tomkins et al. 2011).
	Sensitivity to aversives	Noise sensitivity, gunshot tests, sudden appearance tests	• Steadiness / sureness during gun tests, marking of behavioural postures (Fuchs et al. 2005; Sinn et al. 2010; Slabbert & Odenaal, 1999). • Avoidance reactions during gun tests (Svartberg, 2002; Svartberg, 2005; Svartberg et al. 2005) • Startle reaction to visual and acoustic stimuli (Asher et al. 2013; Sherman et al. 2015; Svartberg, 2002; Svartberg, 2005; Svartberg et al. 2005; Svobodova et al. 2008) • Body posture to unfamiliar stimuli (DeMeester et al. 2008, 2011). • Latency to recover from noise (Batt et al. 2008; Tomkins et al 2011).

Figure 3.3. Categorisation of behaviours assessed within the working dog selection literature to create thematic groups based on core affect.

3.2.4 Analysis

Data obtained in the papers relating to inter-rater, intra-rater (within session), test-retest reliability and predictive validity were pooled at a behavioural trait level (as determined in the first stage of the analysis). This information was then examined to give an indication of the overall quality of selection tests in measuring specific behavioural traits potentially relevant to a variety of working dogs and to evaluate if these traits were predictive of successful performance in the field.

3.3 Results:

The initial literature search, using the terms specified in Figure 3.1, produced 5,300,268 hits, with an additional four references obtained from grey literature searches. After title and abstract screening, records which appeared to match the inclusion criteria (n = 106) were screened for, and excluded based on: duplicates (n = 8), review-articles (n = 18), meta-analysis (n = 1) and lack of public access (n = 13). The remaining 66 references were assessed for full-text eligibility, with 51 exclusions made based on not assessing dog behavioural characteristics relevant to working dogs, not using behavioural tests, or a combination of both these factors (see Figure 3.2). The remaining fifteen papers were fully assessed in accordance with the two stages described in the methods above. In response to the requests for further data from corresponding authors, one author declined to comment as this was being used for future work and another indicated there was no further information. The rest did not respond, or the corresponding author's email was no longer active. The data relating to the articles examined and their classification are summarised in Figure 3.3. A range of behavioural tests and parameters were used to measure positive emotional states, such as body posture during human contact (e.g. Asher Blythe, Roberts *et al*. 2013), time to release toy (Sinn *et al*. 2010) and latency to catch toy (Batt *et al*. 2008). There was also a large range of behaviour tests and parameters used to assess negative emotional states, such as level of aggressive response when provoked (Fuchs *et al*. 2005; Slabbert & Odenaal, 1999), startled (Svartberg, 2002; Svartberg, 2005; Svartberg *et al*. 2005), or approached (Sherman *et al*. 2015) and behaviours during gun tests (Fuchs *et al*. 2005; Sinn *et al*. 2010; Slabbert & Odenaal, 1999).

Details of the assessment of the quality metrics of each trait are given in Figure 3.4 and elaborated on further in the following text.

3.3.1 Inter-rater reliability

There was little reporting of inter-rater reliability statistics (agreement across two or more independent observers). Two papers discussing positive affect behaviours touched upon inter-rater reliability (Asher *et al*. 2013; Sinn *et al*. 2010). However, only one of these papers reported inferential statistics - in the form of significant correlations between raters' scoring behaviours surrounding 'willingness to work', and 'object directed play' at three time points (Sinn et al. 2010). For negative affect behaviours, five papers considered inter-rater reliability (De Meester et al. 2008, 2011; Fuchs *et al*. 2005; Sherman *et al*. 2015; Sinn et al. 2010). Two of these papers reported inferential statistics to support the statement that ratings on sensitivity to aversives, approach-withdrawal and human-directed aggression were reliable across raters (Sherman *et al*. 2015; Sinn *et al*. 2010). With the aim of validating the 'emotional reactivity test' to assess performance in explosive device detection dogs, Sherman *et al*. (2015) reported high correlation coefficients for sensitivity to aversives (minimum of 0.75; visual and acoustic startle, response to remote control car) and moderate coefficients for human-directed aggression (0.6; stranger examination). The authors also reported aggregate scores, which improved correlation statistics considerably (0.96). With the aim of assessing the validity of a behavioural test for use in military working dogs, Sinn *et al*. (2010) reported significant correlations across four raters for sensitivity to aversives (gun sureness) and threat aggression (human directed aggression), separately at each time point and aggregated across two test points.

3.3.2 Intra-rater reliability

Fewer studies reported intra-rater reliability statistics (agreement within a single rater's scores within a session). For positive core affect behaviours, two papers (Asher *et al*. 2013; Sinn *et al.* 2010) mentioned intra-rater reliability, claiming 90% agreement between time points (Asher *et al*. 2013), and good intra-class correlation coefficients (Sinn *et al*. 2010), but there was a lack of reporting of statistical tests to support these claims directly. For negative affect behaviours three papers referred to 'good' intra-rater reliability (Asher *et al.* 2013; De Meester *et al*. 2011; Sinn *et al*. 2010); again there was a lack of statistics presented to support the claims.

3.3.3 Test-retest reliability

Test-retest statistics were reported by three papers discussing positive affect behaviours (Fuchs *et al*. 2005; Sinn *et al.* 2010; Svartberg *et al*. 2005), with mixed findings. Assessing behaviours surrounding willingness to work, Sinn *et al*. (2010) reported significant correlations across and between three time points for some behaviours (object focus, sharpness, and human focus), but less reliable correlations for other behaviours (search focus). With the aim of assessing consistency of behaviours displayed in behavioural tests, Svartberg *et al*. (2005) reported significant correlations and non-significant tests of difference (Friedman's method) across time points, providing stronger statistical evidence for the test-retest reliability of the test used, for behaviours associated with object-directed play (playfulness) and human-directed social behaviours (sociability). Fuchs et al (2005) explored what external factors affect the result of behavioural tests over time, they reported that scores of 'hardness' (object-directed play) did not significantly differ one year later, indicating good test-retest reliability of this trait.

From the papers pertaining to negative affect behaviours, four reported test-retest statistics. One of these papers reported aggregate scores across sub-tests, therefore it is not possible to directly associated the data with specific behavioural aspects, nonetheless this paper reported high coefficients across time points (α = 0.89) (Sherman *et al*. 2015). Three papers reported significant correlations over three time points for behaviours specifically associated with human-directed aggression (Fuchs *et al*. 2008; Sinn *et al*. 2010; Svartberg *et al*. 2005) and approach-withdrawal tendencies (Svartberg *et al*. 2005). However, the two papers (Fuchs *et al*. 2008; Svartberg *et al*. 2005) which computed tests for differences across time points, observed that scores for aggressiveness and curiosity (withdrawal) significantly decreased over the three testing times (Svartberg *et al*. 2005), and that defence drive ratings were significantly different from test 1 to test 2 (Fuchs et al. 2008), suggesting a lack of test-retest reliability. Three papers reported test-retest statistics for behaviours relating specifically to sensitivity to aversives (Fuchs *et al*. 2005; Sinn *et al*. 2010; Svartberg *et al*. 2005). Findings were mixed, with evidence of good (85%) agreement between two test times and no statistically significant differences in rating (Fuchs *et al*. 2005), but also evidence of poor correlations across three time points (Sinn et al. 2010), and significantly different ratings despite evidence of correlations (Svartberg *et al*. 2005).

3.3.4 Predictive validity

Ten of the papers discussing positive affect behaviours reported data pertaining to the predictive validity of the tests used. Behaviours associated with willingness to work predicted success in guide dog training (retrieve response to stimuli: Asher *et al.* 2013; distraction and passive test success: Batt *et al*. 2008) and police/patrol dog certification/efficiency (search focus and sharpness: Sinn *et al*. 2010; retrieve performance at eight-weeks: Slabbert & Odendaal, 1999; decreased scores on the factor 'movement'; Svobodova *et al.* 2008). However, Wilsson and Sundgren (1998) found limited utility in their behavioural test, which assessed behaviours relating to willingness to work, for predicting future service dog performance. Some behaviours associated with human-directed social behaviour were associated with success in guide dog training (stroking response to assessor: Asher *et al*. 2013), better performance in working dog trials (sociability towards strangers: Svartberg, 2002), police dog efficiency tests (factor for movement, incorporating behaviours towards a person: Svobodova *et al*. 2008) and greater cooperation at maturation (Wilsson & Sundgren, 1998). Additionally, scores of sociability (Dog Mentality Assessment) correlated with the dog's social behaviours in the dog's home (Svartberg, 2005), demonstrating convergent validity, and ratings of greeting, cooperation and handling loaded together onto the same 'sociability' factor, highlighting component behaviours associated with human-directed social behaviours (Svartberg & Forkman, 2002).

Behaviours associated with object directed play did not significantly predict success in guide dog training (squirrel-response to stimuli: Asher *et al*. 2013; Latency to catch: Batt *et al*. 2008), or service dog performance (Wilsson & Sundgren, 1998), but did predict success in police dog efficiency tests (attitude to predation, including retrieval and tug of war: Svobodova *et al* 2008) and performance in working dog trails (boldness, related to playfulness: Svartberg, 2002). Furthermore, playfulness (Dog Mentality Assessment) significantly correlated with trainability and human directed play interest behaviours in the home, demonstrating convergent validity (Svartberg, 2005) and ratings of interest in play, grabbing and tug of war loaded on to the factor 'playfulness' (Svartberg & Forkman 2002), indicating component behaviours associated with object-directed play.

Thirteen of the papers reporting negative affect behaviours discussed the predictive validity of the tests used. Behaviours associated with human-directed aggression at 6-9 months predicted police dog efficiency (Slabbert & Odendaal, 1999), but no other reports of human-directed aggression predicting future working performance were mentioned. There is evidence of some convergent validity for this behavioural characteristic, with posture score in 'contact with strangers' correlating with 'fear of strangers' in Canine Behavioural Assessment and Research Questionnaire (CBARQ) (De Meester *et al*. 2008), and scores of sociability in the Dog Mentality Assessment correlating with stranger-directed aggression (Svartberg, 2005). Additionally, aggression in: distance play, sudden appearance, and the ghost tests loading onto the factor 'aggression', indicating these tests may be useful for assessing human-directed aggression behaviours (Svartberg & Forkman 2002). In contrast, De Meester *et al.* (2011) reported that the Socially Acceptable Behaviour (SAB) test was not useful for predicting fight or flight behaviours, and therefore may not prove to be a valid test to measure human-directed aggression in working dogs. With regards to what factors predict the behaviours associated with human-directed aggression, Fuchs *et al.* (2005) reported that early and frequent contact with school age children predicted a more desired "defence drive"; although not directly relevant to evaluation of behavioural tests, this is nonetheless a relevant finding when considering predictive relationships associated with dog behavioural tests.

There was conflicting evidence as to whether behaviours associated with sensitivity to aversives predicted success in guide dog training. Reports of latency to recover from noise predicted guide dog success when tested at 12 and 14 months (Batt *et al.* 2008) and latency to sit during passive and noise tests predicted success when tested at 13-17 months (Tomson *et al*. 2011). However, behavioural responses (e.g. shaking) to noise at 6-8 weeks did not (Asher *et al*. 2013). Similarly, gunshot sensitivity did not predict adult police dog efficiency whereas startle test responses were higher in those who became police dogs than those who did not (Slabbert & Odendaal 1999). In contrast, a more positive (less fearful) response at 7 weeks of age, to noise predicted a lower probability of passing police dog training (Svobodova *et al*. 2008), whereas dogs who scored high boldness (related to fear) performed significantly better in working dog trials (Svartberg, 2002). There was a lack of convergent validity for behaviours associated with sensitivity to aversives in relation to scores on the CBARQ (De Meester *et al.* 2008) and the SAB (De Meester *et al*. 2011). Startle reaction, and other behaviours associated with sensitivity to aversives were found to load on to the factor aggression (Svartberg & Forkman, 2002), indicating there may be a relationship between these two behavioural characteristics, which should be considered in the development of future behavioural tests. With regards to what factors predict the behaviours associated with sensitivity to aversives, Fuchs *et al.* (2005) reported that young dog training predicted a more desired nerve stability.

There was little evidence of applying withdrawal behaviours to predicting working dog success. One study explored these tendencies in relation to guide dog performance, but failed to find a signification relationship (Tomkins *et al*. 2011). Nonetheless, work by Svartberg and his colleagues suggests that approach-withdrawal tendencies associated with 'curiosity' displayed in behavioural tests relates to everyday behaviours including social and non-social fear (Svartberg, 2005; Svartberg & Forkman, 2002) indicating some possible predictive validity for tests which assess this characteristic.

Behavioural Characteristic & References	Inter-rater	Intra-rater	Test-retest	Predictive validity
Willingness to work				
Asher et al 2013	Scores were replicated in ~90% of cases (no further statistics presented)	Scores were replicated in ~90% of cases (no further statistics presented)	X	Retrieve-response to stimuli and ramp tests predicted success in guide dog training (Chi Square tests, $p < 0.05$).
Batt et al 2008	X	X	X	Distraction test and passive test predicted guide dog success (Logistic regression model).
Sinn et al 2010	For search activity, search stamina and attention transfer (relevant to willingness to work) inter-rater scores across the three time points significantly correlated ($p<.01$).	Stated close agreement; intra-class correlation coefficient, but no supporting statistics presented.	Spearman's correlations were significant for comparisons of object focus, sharpness, human focus and search focus between time 1-2, time 1-3, time 2-3 ($ps < 0.01$), with the exception of search focus between time 2-3 ($p=0.32$)	Search focus and sharpness improved prediction scores of certification outcome in patrol dogs ($p < 0.01$; Wald Chi-square).
Slabbert & Odendaal 1999	X	X	X	Retrieve performance (8 & 12 weeks) predicted adult police dog efficiency ($p < 0.01$; statistical test unclear).
Svobodova et al 2008	X	X	X	Negotiating obstacles loaded on to the factor for movement. Lower scores on 'movement' predicted greater chance of passing police dog test.
Wilsson & Sundgren 1998	X	X	X	Generally, not useful for predicting service dog work.
Human directed social behaviour				
Asher et al 2013	Scores were replicated in ~90% of cases (no further statistics presented)	Scores were replicated in ~90% of cases (no further statistics presented)	X	Stroking response to assessor was associated with success in guide dog training (Chi Square tests, $p < 0.05$).
Svartberg 2005	X	X	X	Sociability (Dog Mentality

				Assessment) is correlated with corresponding behaviours in the dog's home (lower stranger directed aggression and fear, greater stranger-directed interest; *ps* <0.01).
Svartberg 2002	X	X	X	Dogs who scored high on boldness (related to sociability towards strangers) scored significantly higher in working dog trials than those who were scored medium or low in boldness (Kruskal Wallis ANOVA, $p < 0.05$)
Svartberg & Forkman 2002	X	X	X	Ratings of greeting, cooperation and handling loaded (>0.5) on to the factor 'sociability'.
Svartberg et al 2005	X	X	Scores for the trait sociability significantly correlated over 3-test points (r> 0.57). Friedman's tests showed values were not significantly across time points (p=0.41).	X
Svobodova et al 2008	X	X	X	Decreased scores on 'factor for movement' (including behaviour towards a person) predicted increased probability of passing police efficiency test (p = 0.02).
Wilsson & Sundgren 1998	X	X	X	Reaction to a stranger at eight-weeks predicted co-operation at maturity in German Shepherd's.
Object directed play tendencies				
Asher et al 2013	Scores were replicated in ~90% of cases (no further statistics presented)	Scores were replicated in ~90% of cases (no further	X	Squirrel-response to stimuli (i.e. playing with toy) did not predict success in guide dog training

		statistics presented)		(Chi Square tests, $p > 0.05$).
Batt et al 2008	X	X	X	Latency to catch did not significantly predict guide dog success (variable removed from logistic regression model).
Fuchs et al 2005	Quality of assessment between raters was sufficiently identical for all judges although significant differences were shown in a very large data set involving about 50–60 judges during several decades.	X	Scores of hardness (rough toy play) did not significantly differ across test times (1 year apart).	X
Sinn et al 2010	Inter-rater scores for interest in objects significantly correlated, across 3 time points.	Stated close agreement; intra-class correlation coefficient, but no supporting statistics presented.	Inter-rater scores for interest in objects significantly correlated, across 3 time points.	X
Svartberg 2005	X	X	X	Playfulness (Dog Mentality Assessment) is significantly correlated with trainability and human directed play interest behaviours in the home (*ps* < 0.01).
Svartberg 2002	X	X	X	Dogs who scored high on boldness (related to playfulness) scored significantly higher in working dog trials than those who were scored medium or low in boldness (Kruskal Wallis ANOVA, $p < 0.05$).
Svartberg & Forkman 2002	X	X	X	Ratings of interest in play, grabbing and tug of war, interest in play and grabbing loaded (>0.6) on to the factor ‘playfulness’.

Svartberg et al 2005	X	X	Scores for the trait playfulness significantly correlated over 3-test points (r> 0.76). Friedman's tests showed values were not significantly across time points (*p*=0.95).	X
Svobodova et al 2008	X	X	X	Increased scores on 'factor for attitude to predation' (including retrieval and tug of war) significantly predicted increased probability of passing police efficiency test ($p < 0.01$).
Wilsson & Sundgren 1998	X	X	X	Not useful for predicting service dog work. Tug of war at 8 weeks did not correlate with traits at an older age.
Human directed aggressive behaviours				
DeMeester et al 2008	If differences in their judgments, images were analysed at a lower speed, and discussion held until agreement	X	X	Posture score in 'contact with strangers' correlated with 'fear of strangers' in Canine Behavioural Assessment and Research (CBAR) questionnaire.
DeMeester et al 2011	Reports of inter-rater reliability, but no associated data is made available.	Reports of intra-rater reliability, but no associated data is made available.		Not possible to predict flight or bite behaviour based Socially Acceptable Behaviour (SAB) test.
Fuchs et al 2005	Quality of assessment between raters sufficiently identical for all judges although significant differences were shown in a very large data set involving about 50–60 judges during several decades.	X	Defence drive ratings were significantly different from test 1-test 2 ($p < 0.01$).	Early and frequent contact with school age children predicted a more desired defence drive ($p < 0.01$; logistic regression analysis)
Sinn et al 2010	Inter-rater scores for threat aggression, non-threat bite quality, threat bite	Stated close agreement; intra-class correlation	Threat aggression, non-threat bite quality, threat bite quality significantly	X

	quality significantly correlated, across 3 time points (p<.01)	coefficient, but no supporting statistics presented.	correlated across 3 time points (p<.01)	
Sherman et al 2015	Inter-class correlation statistics for aggression during stranger approach were moderate (0.65)	X	Aggregate scores across 17 behavioural tests significantly correlated (p < 0.01) (α = 0.89).	X
Slabbert & Odendaal 1999	X	X	X	Aggression (6 & 9 months) predicted adult police dog efficiency.
Svartberg 2005	X	X	X	Sociability (Dog Mentality Assessment) was significantly correlated with stranger-directed aggression (ps < 0.01)
Svartberg & Forkman 2002	X	X	X	Aggression in distance play, sudden appearance, ghost tests loaded onto the factor 'aggression' (>0.4)
Svartberg et al 2005	X	X	Scores for trait aggressiveness significantly correlated over 3-test points (r> 0.68). However, Friedman's tests showed values significantly decreased across time points (p<0.01).	X
Wilsson & Sundgren 1998	X	X	X	Generally, not useful for predicting service dog work. Objects visited at 8 weeks correlated with defence drive when older (3-6 yrs) (p = 0.02).
Withdrawal tendencies				
Sherman et al 2015	Unusual stranger reaction (i.e. approach) revealed high inter-rater coefficient scores (0.88)	X	Aggregate scores across 17 behavioural tests significantly correlated (p < 0.01) (α = 0.89).	X
Svartberg 2005	X	X	X	Curiosity significantly correlated with stranger directed fear, non-social fear

				and human-directed play interest.
Svartberg & Forkman 2002	X	X	X	Exploration in sudden appearance test, metallic noise and ghost test loaded onto the factor 'Curiosity and Fearlessness'.
Svartberg et al 2005	X	X	Scores for curiosity were significantly different (increased) over test times (p = 0.009).	X
Tomkins et al 2011	X	X	X	Approach behaviours in the sudden appearance test did not predict guide dog success.
Sensitivity to aversives				
Asher et al 2013	Scores were replicated in ~90% of cases (no further statistics presented)	Scores were replicated in ~90% of cases (no further statistics presented)	X	Noise response did not predict success in guide dog training (Chi Square tests, $p >$ 0.05)
Batt et al 2008	X	X	X	Latency to recover from noise significantly predicted guide dog success at 12 months on old (4/5 tests) and 14 months (2/5 test times).
DeMeester et al 2008	If differences in judgments, images were analysed at a lower speed, and discussion held until agreement.	X	X	For sub-tests where the dog was exposed to unfamiliar sounds or visual stimuli, there was no correlation with 'non-social behaviour' on CBARQ.
DeMeester et al 2011	Reports of inter-rater reliability, but no associated data is made available.	Reports of intra-rater reliability, but no associated data is made available.		Not possible to predict flight or bite behaviour based on Socially Acceptable Behaviour (SAB) test.
Fuchs et al 2005	Quality of assessment between raters sufficiently identical for judges although significant differences in a large data set involving	X	Nerve stability ratings 85% agreement over testing times – no statistically significant differences	Young dog training predicted a more desired nerve stability (p = 0.01; logistic regression analysis).

	about 50–60 judges over several decades		between time 1-time 2 ($p > 0.05$)	
Sinn et al 2010	Inter-rater scores on gun sureness considerably varied over the three time points, but correlation coefficients were statistically significant.		Gun sureness did not correlate between time 1-2, or time 1-3, only between time 2-3 ($p < 0.01$)	X
Sherman et al 2015	Inter-rater scores for tests relating to sensitivity to aversives moderate to high (visual startle = 0.7; acoustic startle = 0.8; remote control car = 0.7).	X	Aggregate scores across 17 behavioural tests significantly correlated ($p < 0.01$) ($\alpha = 0.89$).	X
Slabbert & Odendaal 1999	X	X	X	Gunshot sensitivity did not predict adult police dog efficiency ($p > 0.05$, statistical test unclear). Higher performance on the startle test at 12 and 16 weeks was observed in those who became police dogs, compared to those who did not ($p < 0.01$).
Svartberg 2005	X	X	X	Curiosity/fearlessness (Dog Mentality Assessment) was significantly correlated with stranger-directed fear, non-social fear and human-directed play interest ($ps < 0.05$).
Svartberg 2002	X	X	X	Dogs who scored high on boldness (related to fearlessness) scored significantly higher in working dog trials than those who were scored medium or low in boldness (Kruskal Wallis ANOVA, $p < 0.05$).

Svartberg & Forkman 2002				Startle reaction, exploration and remaining avoidance in the in the sudden appearance and metallic noise test, along with exploration in the ghost test, loaded onto the 'aggression' factor (>0.4).
Svartberg et al 2005	X	X	Scores for the trait curiosity/ fearlessness significantly correlated over 3-test points (r> 0.58). However, Friedman's tests showed values significantly increased across time points (*p*<0.01).	X
Svobodova et al 2008	X	X	X	Higher scores on 'factor for responding to noise' (including, response to distracting stimuli caused by a shovel, response to a distracting noise when left in a room and response to loud distracting stimuli) predicted lower success probabilities for passing police test (*p* = 0.02).
Tomkins et al 2011	X	X	X	Latency to sit in the passive and noise tests predicted guide dog success (*p* = 0.02). The presence of panting (*p* = 0.02) and licking (p = 0.005) in the dog distraction test reduced probability of passing the test.
Wilsson & Sundgren 1998	X	X	X	Not useful for predicting service dog work.

Figure 3.4: Reported information for reliability and validity for behavioural characteristics relevant to working dogs from the papers selected for the review.

3.4 Discussion:

With the aim of identifying the reliability and validity of tests of potential value for predicting the performance of working dogs, standard search procedures were used to identify available statistics of the behavioural tests described in the literature. It was evident from reviewing the papers that a large range of tests are used (e.g. gun fire, sudden appearance test, obstacle courses, stranger approach, toy/Kong tests) with a range of parameters to indicate performance (e.g. subjective rating scores of body postures, scores of vocalisations, time taken to achieve target). It was observed that some of these tests not only assessed similar traits, but also labelled similar traits using different terminologies. These behavioural tests were therefore grouped according to the putative traits they assessed relating to positive affect (willingness to work, object directed play, human directed social behaviour) and negative affect (human directed aggression, sensitivity to aversives, withdrawal tendencies). If accurate, direct, comparisons are to be made of the validity of behavioural tests and make inferences on the importance of specific traits for working dog performance researchers need to be aware of the importance of using consistent terminologies.

3.4.1 Reliability

For behaviours associated with positive affect, only one study reported significant correlation between raters (Sinn *et al*. 2010), but this failed to report Cohen's Kappa (Cohen, 1960), or alternative coefficient statistics (McHugh, 2012). For behaviours associated with negative affect behaviours two papers were identified as providing supportive statistics (Sinn *et al*. 2010; Sherman *et al*. 2015). No papers reported Intra-class Correlation Coefficients (ICC) for within rater reliability assessment (Yen & Lo, 2002). This leaves considerable doubt over any claim that the data obtained from these tests is objective, since without showing some consistency between or within observers, this cannot be known.

Evaluation of test-retest reliability results further highlight the importance of reporting statistical test results, with tests reporting significant correlations over testing times, but also significant differences between values over time (e.g. Fuchs et al. 2005 Svartberg et al. 2005). This is an important point since it is possible to have good correlation without repeatable results (i.e. the intercept of the correlation is not through zero), and any consistent difference between tests needs to be known so it can potentially be corrected for. Thus correlation alone is not sufficient to show that results are completely reliable. It should, however, be noted that test-retest reliability is likely to be affected by the study design as well as the reliability of the behaviour test. For instance, Fuchs *et al*. (2005) had an average of a 1-year delay between test and retest, whereas Svartberg *et al.* (2005) had an average delay of 35 days. Nonetheless, there was some evidence for test-retest reliability for both negative affect behaviours (sensitivity to aversives; Fuchs et al. 2005; Sherman *et al*. 2015) and positive affect behaviours (object-directed play and human-directed social behaviours; Svartberg et al. 2005), as supported by tests of correlation and tests of difference. This suggest that both the trait and method of assessment are probably robust. Test-retest reliability may be particularly important when considering the difficult task of examining the difference between a behaviour and a trait (Wilsson & Sinn, 2012). Traits are typically inferred from a behaviour which has been observed across situations (i.e. high test-retest reliability), whereas specific behaviours may disappear over time due to changes such as behavioural habituation, rather that unreliability of the test per se, highlighting the importance of considering other features of the test, such as predictive validity.

3.4.2 Predictive Validity

Predictive validity was the most frequently reported metric. Positive affect behaviours related to performance in guide dog roles (willingness to work, human-directed social behaviour; Asher et al. 2013; Batt et al. 2008), police dog work (willingness to work, human-directed social behaviour and object-directed play; Svobodova et al 2008), and in working dog trials (human-directed social behaviour, object-directed play; Svartberg, 2002). With the exception of one paper (Svartberg, 2002) these conclusions are based on the results of logistic regression analyses.

There was less consensus across the papers on whether negative affect behaviours predicted working dog success. Only one report indicated that human-directed aggression predicted police dog efficiency (Slabbert & Odendaal, 1999), and there were conflicting results with regard to sensitivity to aversives. Indeed, whereas gunshot sensitivity did not predict police dog success, startle response did (Slabbert & Odendaal, 1999), and the initial startle may be a better predictor of general autonomic sensitivity, since the response beyond this will depend on higher level appraisal of coping ability. Furthermore a more positive response to noise at 7-weeks old was associated with a lower likelihood of passing police dog certification (Svobodova *et al.* 2008). This highlights the importance of consistency of test characteristics and requirements when assessing behavioural traits. Similarly, with regard to guide dogs, latency to recover from an aversive stimulus did not predict later guide dog success (Batt *et al*. 2008), whereas behavioural response to an aversive stimulus did (Asher et al. 2013), emphasising the importance of using consistent parameters when comparing performance across tests. However, it could also be that these contrasting results reflect age-related developmental differences in the dogs, with Asher *et al* (2013) working with younger dogs (6-8 weeks) than Batt *et al* (2008) (testing started at 6 months). Nonetheless, this would stand in contrast to reports which suggest that important guide dog traits can be measured more reliably in older puppies (>14 months; Batt *et al.* 2008). It should also be considered that the simple pass/fail criteria used to determine guide dog success ignores the disparity of outcomes which may be associated with successful performance as a guide dog, which limits the test specificity and accuracy.

3.4.3 Convergent Validity

Convergent validity i.e. the extent to which the test covaries with other tests that should theoretically be related, was reported in two of the papers (Foyer *et al*, 2016. Sherman *et al*, 2015). Both found good convergent validity between behavioural measurements and salivary cortisol measures; finding an increase in levels of cortisol following the behavioural tests they administered and dogs that scored higher for emotionality (Foyer *et al*, 2016) or showing greatest emotional reactivity (Sherman *et al*, 2015). This was further validated by convergence with an open field anxiety test by Sherman *et al*, 2015, where dogs with greatest emotional reactivity showed greater increases in anxiety when exposed to fear eliciting stimuli. Foyer *et al*, 2016, also demonstrated further convergence through an association between emotionality and passing the selection test (convergent predictive validity). Both of these papers focus mostly on 'sensitivity to aversives' in this regard, suggesting that this trait may be reliably measurable and a useful construct for assessing working dogs. Fear responses are often more rigid, and have also been found to have higher heritability values in dogs than positive behavioural traits (Wilsson & Sundragen, 1998. Ruefenacht *et al*, 2002). Aspects of positive activation may be less easy to predict and more complex at a number of levels. In questionnaire measures such as the PANAS positive activation is further broken down into components of Energy & Interest, Excitement, and Persistence, perhaps making it more complex and harder to measure than Negative Activation (Taylor & Mills, 2006).

3.4.4 Inconsistency of Terminology

The inconsistent terminology used to describe traits being inferred from the behaviours they were measuring, posed certain challenges. It was necessary to group the measures of behaviour from the tests into categories that related more broadly to a common trait. This enabled an assessment of a wider range of quality metrics for a higher level trait that might be relevant to working dogs. The need for reclassification creates potential difficulties with respect to objectively comparing tests, and questions the value of the tests being implemented as they were intended by different user groups. Working dog user groups would benefit from agreeing and defining standardised terminology for the traits they feel are important in terms of working dog success or withdrawal to enable such comparisons. The lack of consensus in the literature as to what these terms objectively mean, and the interchangeable use of terminology also make it difficult to keep track of knowledge and progress in the field (Jones & Gosling, 2005).

Behavioural characteristics relevant to working dogs are frequently examined in the literature and while similar traits are often assessed using similar methods, it is often with different descriptions or labels (Jones & Gosling, 2005). For example, many tests appear to examine temperament on a shyness boldness continuum (Svartberg, 2002), although some tests will call this confidence and fearfulness (De Meester *et al*, 2011). It is important to understand that while we infer traits based on measuring behaviour, what people label a trait, even the same trait, is based on different behavioural evidence.

3.4.5 Temperament and Working Dogs

Regardless of the nature of a dog's job, temperament is important for a dog to fulfil its role and certain traits are relatively consistently referred to and putatively measured in the literature despite the differences in working dog roles. The most commonly measured trait related to sensitivity to aversives, with the majority of behaviour tests aiming to examine fearful type responses to potentially threatening stimuli. This suggests that in working dog selection, testers probably want to know about how sensitive the dog is to stressful stimuli, as logically they are looking for dogs that are able to cope when put into situations which are stressful or potentially could elicit fear. This shows a need to understand how potential working dogs respond to negative qualities in the environment, and how they recover from these. Different temperament traits are important to different extents depending on the role of the dog, for example aggressivity may be desirable in a military working dog, but undesirable in a guide dog, but knowing about it is important to both types of working dog. It is therefore important to identify the optimal behavioural phenotype for certain roles (Wilsson & Sinn, 2012), to enable better selection of dogs for working positions which could reduce dropout rates, increase success of certification and also improve the dog's welfare, as certain dogs may not be psychologically robust enough for certain working roles.

Testing for psychological characteristics which reflect temperament in dogs does not need to be limited to behavioural measures. While the focus of this review was aimed at gathering information about the available behavioural tests as these are the most common form of assessing working dogs, it is important to note that other measurement methods could be useful. Physiological measures can be used to examine underlying physiology and/or physiological responses to situations and can look to correlate this with what is observed in behaviour tests (e.g. Sherman *et al*, 2015 who analysed salivary cortisol alongside the Emotional Reactivity Test). Both behavioural tests and questionnaire measures for trait impulsivity have been shown to have convergent validity with urinary metabolites of serotonin and dopamine (Wright et al, 2012). This predicted convergence gives confidence that the methods used to assess this trait are robust. In some situations it may not be possible to use a behaviour assessment, for example if a dog is physically impaired or in pain this could impact on their performance. A lack of space, time, or access to the dogs needing to be assessed could also prevent the use of both physiological and behavioural measures. Questionnaire data based on experience with the dog over a prolonged time can be used to get behavioural information about an individual from the owner, handler or trainer to identify the temperament or ability of a dog. This is the principle used in both the PANAS (Sheppard & Mills, 2002) and the DIAS (Wright *et al*, 2011) which have been validated to varying degrees but both may be of relevance to describing temperament in working dogs. These scales provide a potential basis for objectively describing broader factors than behavioural performance in supposedly related tasks that may be associated with success or withdrawal. While questionnaires can be used to determine how a dog is likely to respond in a given set of circumstances to build up a picture of the dog's behaviour, this can be problematic as it requires someone to have sufficient knowledge of the dog to complete the questionnaire, and also that the questionnaire is reliable and valid for use in dogs of the age being assessed. It also relies on receiving accurate and honest information from those completing it, but particularly within the working dog sector, there may be concerns over the reliability of such information being provided by individuals who may have an interest in selling dogs of a certain type, given the high value of working stock. While behavioural tests are usually the favoured form of temperament assessment, a combination of the 3 measures (behavioural, physiological and questionnaire) brings convergent validity to the process and strengthens any conclusions. Due to their reliability as measures of temperament traits, the PANAS (Sheppard & Mills, 2002) and DIAS (Wright *et al*, 2011) are potentially a good starting point for further development of reliable and valid behaviour assessments for profiling successful working dogs, as well as those that have been withdrawn, with a view to use this information to inform the selection process.

This review was limited by the restricted availability of original data sources that might have helped assess the reliability and validity of the tests. We suggest that such data be made available as a matter of routine either within publications or within accessible electronic repositories, if some restriction is required. Although the search strategy used within this review means other relevant publications likely exist, the papers presented in this review provide a reasonable representation of the types of tests in use for which data are available, and if anything overestimate the reporting of quality metrics, since they tend to be well-cited.

3.5 Conclusion:

This review highlights issues such as inconsistent terminology and a variation in methods in the way that working dogs are being selected that are probably impacting on the quality of animals being used. There is a need to ensure that behavioural characteristics are being tested in a reliable way that is predictive of later field performance which has also been accurately defined. There needs to be consistency with the way that working dog traits are described and measured to ensure that dogs are being assessed appropriately for working roles. Tests should ideally use a combination of behavioural, physiological and questionnaire measures to ensure internal consistency of traits being measured and provide convergent validity for the constructs of interest. The process used to gather this information needs to be quick, simple and effective. There is a need for uniformity in methods of testing temperament (Deidrich & Giffroy, 2006). Simple, quick, practical tests to assess the temperament of an animal would be valuable tools for those working with and involved in the care of animals (Seaman *et al*, 2002). As highlighted, testing core affect as well as trait impulsivity may be useful as an additional assessment measure in working dogs to help further inform selection of traits that have been highlighted within this chapter as relevant.

Chapter 4: Developing a Simplified Behaviour Test for Impulsivity in dogs

Synopsis

The previous chapter highlighted room for additional measures that could be utilised in identifying temperament traits of relevance to working dogs. There are reliable and valid measures for assessing temperament trait impulsivity in pet dogs, which may be of relevance to working dog performance as discussed in Chapter 1, making this a starting point for assessments provided tests can be sufficiently simplified for use in working dogs.

In domestic dogs trait impulsivity can be measured psychometrically using the Dog Impulsivity Assessment Scale (DIAS) and experimentally using a temporal discounting paradigm which requires substantial training. A Spatial Discounting Task (SDT) was developed as an alternative experimental method to assess impulsivity, and evaluated performance in adult (2-10 years, N=24) and younger (2-9 months, N=24) dogs. The test was modified for field use with fewer controls (Simplified Spatial Discounting Task (SDTs), N=13). Convergent validity with the SDT and DIAS Overall Questionnaire Scores (OQS) and stability over time (4-6 weeks) in the two age groups was determined. 96% of dogs recruited reached criterion for testing. A significant positive relationship was found between Maximum Distance Travelled (MDT) in the SDT and OQS in adult dogs ($r=0.46$, $p=0.028$), with good test-retest reliability evident for both ($p<0.001$, N=12). In young dogs there was good test-retest reliability for OQS ($p=0.023$, N=12), but no significant relationship was found between OQS and MDT, test-retest reliability for MDT in young dogs was poor. In the SDTs, 100% of dogs recruited met criterion for testing and there was a significant relationship between MDT and OQS ($r=0.61$, $p=0.027$). The SDT appears to be a useful method for measuring impulsivity in adult dogs with wide applicability.

4.1 Introduction

Impulsivity has been defined as the extent to which an individual evaluates the consequences of their behaviour (Peremans *et al*, 2002), and is expressed in a range of behaviours linked with inhibitory control (Wright *et al*, 2012). Higher levels of impulsivity have been linked to poorer or less accurate choices (Puumala *et al*, 1998) suggesting that this aspect of temperament is potentially an important predictor of performance in complex environments, such as those experienced by working dogs. In addition the trait is thought to be linked to behaviour problems relating to a limited or lack of self-control and an inability to tolerate frustration (Wright *et al*, 2011), again, potentially of importance in the performance of the working dog in the field.

4.1.1 Measuring Impulsivity in Dogs

Temporal discounting i.e. the tendency for rewards to become less valuable the further away they appear to be, has been extensively examined across a number of species including rats (Reynolds *et al*, 2002. Green *et al*, 2004) pigeons (Green *et al*, 2004), dogs (Wright *et al*, 2012) and humans (Dittmar & Bond, 2010). Temporal discounting paradigms have been used to examine cognitive processes linked to decision making in terms of goal directed behaviour, addiction behaviours (Bickel et al, 2007), and socially important behaviours (Critchfield & Kollins, 2001), some of which can relate to impulsivity. Indeed temporal discounting has also been used to examine temperament (Ostazewski, 1996), including impulsivity in rats (Perry *et al*, 2007), pigeons (Ainslie, 1974 & Wolff *et al*, 2002), primates (Tobin et al, 1996. Evans & Beran, 2007. Rosati *et al*, 2007) and humans (Ostazewski, 1996. Bickel et al, 2007). ,

The Dog Impulsivity Assessment Scale (DIAS) is a validated questionnaire which assesses impulsivity in dogs. It consists of one overarching factor, Overall Questionnaire Score (OQS) which gives a general overview of the level of impulsivity in the dog. This OQS is then broken down into three component factors; Behavioural Regulation (Factor 1 (F1), covers the dogs ability to show self-control in its own behaviour), Aggression and Response to Novelty (Factor 2 (F2), covers fearful behaviour in response to potentially threatening or novel stimuli), as well as Responsiveness (Factor 3 (F3), covers how responsive the dog is to the environment and training) (Wright *et al*, 2011). The long term stability of impulsivity has also been examined and found to be stable up to 6 years post initial assessment using the DIAS (Reimer et al, 2014).

As part of the validation of the DIAS, its convergent validity with an operant learning task centred around a temporal discounting paradigm (delayed reward choice task (DRCT), which examines maximum time delay tolerated for a larger reward when an immediate smaller reward is also available, (Wolff *et al*, 2002)) has been assessed (Wright *et al*, 2012). The maximum delay tolerated before switching was found to correlate with impulsivity (more impulsive individuals tolerated shorter delay) (Wright *et al*, 2012). However, this test involves considerable training, with dogs reported to take up to four, 60 minute sessions to reach the criterion required before assessment can begin. A 50% drop out rate has also been reported during this process (Wright *et al*, 2012). These factors limit the utility of this test to assess impulsivity in animals. In addition, although the test is validated against the questionnaire and therefore the questionnaire could be used in its place, the questionnaire is limited to use in dogs with a known history (where there is a consistent caregiver who can answer the questions for the dog over a range of contexts). This means that other means of assessing impulsivity in dogs of unknown history, for example dogs being acquired from breeders as working dogs, would be beneficial.

4.1.2 Spatial discounting as a tool to assess impulsivity

Spatial discounting potentially offers an alternative method to temporal discounting for the assessment of impulsivity. It assesses the animal's perception of reward value as distance to it changes (Stevens *et al*, 2005). The spatial discounting paradigm typically involves assessing the value of reward while gradually increasing the distance an animal is required to travel for a larger reward compared to a smaller one and determining the distance at which it will no longer travel for the larger reward and so switches to the smaller, closer reward. Spatial discounting paradigms are of course not completely independent of temporal choice, as an increase in distance also creates an increase in time taken to reach the larger reward. Similar to temporal discounting, spatial discounting has been used in a variety of ways to help understand decision making behaviour. It has been used in primates to help understand goal directed behaviour relating to foraging patterns (Krali & Sampson, 2012. Stevens *et al*, 2005), to examine evaluative behaviours in rats (Papale *et al*, 2012), and to examine the differing impact of food and social reward on decision making behaviour in guppies (Mulhoff et al, 2011). To our knowledge spatial discounting has not been used to assess individual differences in delay tolerance as a potential measure of impulsivity.

A major potential advantage of a spatial discounting task over a temporal delay paradigm is that no specific training is necessarily required (whereas in the temporal tests dogs first had to learn a lever press task before testing), giving the resulting test the potential to be applied more quickly, with a greater number of animals qualifying for assessment.

4.1.3 Aims

The aim of this study was to develop a spatial discounting task suitable for dogs of all ages and determine its validity as a measure of impulsivity using DIAS scores in adult dogs. Secondarily the convergent validity of the spatial discounting task and DIAS was examined in young dogs. Finally the spatial discounting task was adapted to make it simpler to use in a field setting with fewer controls and examined its convergent validity with DIAS. Retesting was used in the adult dogs and young dogs to determine re-test reliability of the results.

Two variants of a runway task were used to examine the distance at which a large reward was discounted by dogs, in favour of a smaller closer reward. This measure from each variant was then examined against overall questionnaire score (OQS) for impulsivity to establish the preferred form of runway method as a valid behaviour test for assessing impulsivity in dogs and puppies. DIAS scores and spatial discounting task performance in dogs under 9 months of age were then examined to investigate the development of the trait over time within individuals.

4.2 Methods:

4.2.1General principles of the spatial discounting task (SDT) used:

In the spatial discounting task (SDT), a variation of the spatial discounting paradigm, dogs were presented with two trays, one containing 3 pieces of food, and the other containing 1 piece of food. Training involved ensuring the dog can discriminate between where the large and small rewards are located. Once this has been established, testing involved a series of trials in which the dog was allowed to choose between the two trays. The dog naturally chooses the tray with the larger reward initially, but this tray is gradually moved further away from the dog, so it must travel further to get the reward over time. The goal of a spatial discounting task is to establish the distance the dog will travel for the large reward before switching to the smaller but closer reward, determined by a predetermined number of consecutive choices of this smaller reward. At this point the maximum distance travelled to eat from the large tray was recorded. See figure 4.1 and 4.2.

In the version used within this study the trays were of different colour to potentially ease discrimination (Neitz *et al*, 1989), and the larger reward tray moved 25cm at a time. Further details of the specific methods and how they were adapted is given in relation to each study below.

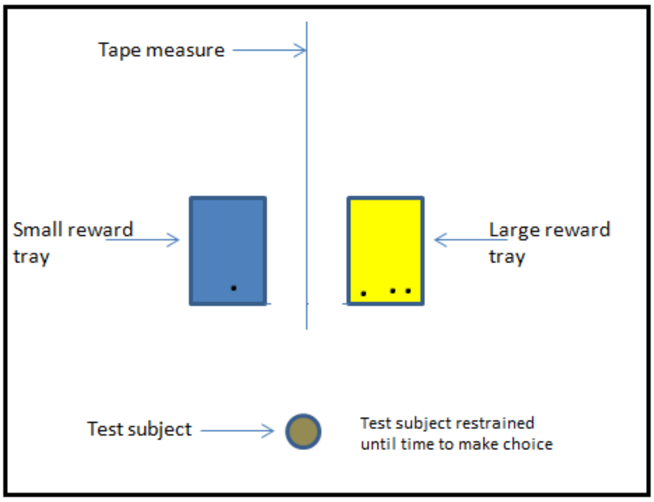


Figure 4.1. Experimental set up for the spatial discounting task (SDT) illustrating the positioning of the equipment and the test set up.

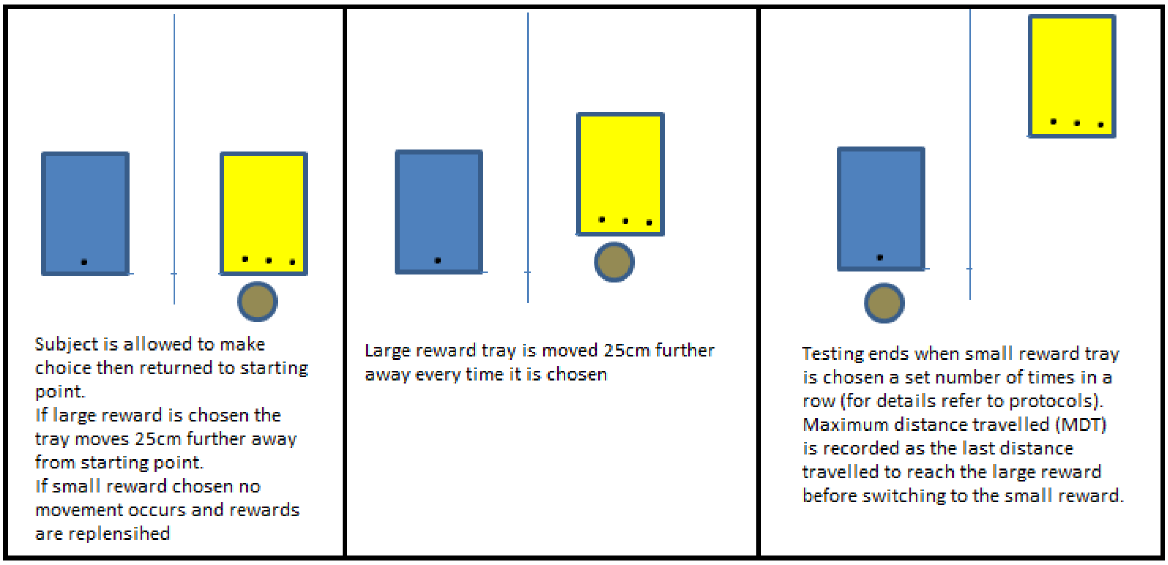


Figure 4.2. Testing procedure for the spatial discounting task (SDT)

4.2.2 Psychometric Profiling

Owners (each dogs primary caregiver, defined as the person in the home who spends most time with the dog on a daily basis) completed The Dog Impulsivity Assessment Scale (DIAS) (Wright et al 2011) to enable impulsivity scores and related components to be calculated. The questionnaire took a maximum of ten minutes to complete and the repeat was 4-6 weeks later. For dogs under 9 months of age which completed the spatial discounting task on two separate occasions, the questionnaire was also completed twice to examine the test retest reliability of the DIAS in younger dogs. From the DIAS, Overall Questionnaire Score (OQS), Factor 1 Behavioural Regulation (F1), Factor 2 Aggression and Response to Novelty (F2), and Factor 3 Responsiveness (F3) scores were all calculated as measures of aspects of impulsivity.

4.2.3 Laboratory Study 1: Spatial Discounting Task (SDT) with adult dogs

4.2.3.1 Training & Testing Dogs

4.2.3.1.1 Subjects

25 adult dogs (with no known existing medical problems relating to sight, olfaction or mobility) over 2 years old were recruited from the general public, 24 of which reached criterion for data inclusion. Half of the subjects (12 adults) were re tested 4-6 weeks after initial testing. For demographics information of the dogs in the study see Figure 4.3.

Age (months)	Sex	Breed
24	M	Labrador
36	M	Crossbreed
96	M	Springer Spaniel
48	F	Miniature Schnauzer
36	F	Sheltie
42	M	Crossbreed
76	F	Border Collie
28	M	Border Collie
96	F	Labrador
61	F	Crossbreed
48	F	Border Collie
86	M	Rottweiler
42	F	Sheltie
50	M	Jack Russel Terrier
52	M	Golden Retriever
102	M	Labrador
26	F	Springer Spaniel
38	F	Jack Russel Terrier
41	M	Crossbreed
118	F	Crossbreed
108	M	Springer Spaniel
62	M	Labrador
71	M	Crossbreed
84	M	Crossbreed

Figure 4.3. Demographic information for dogs included in the Spatial Discounting Task with Adult Dogs.

The tester and assistant were present in the room during testing, owners waited outside the room and were able to watch the testing via a two way mirror, so they could observe without the dog being able to see them.

4.2.3.1.2 *Equipment:*

Two equal sized trays (length 36cm x width 25cm x depth 6cm), one blue, one yellow, were used, with the larger reward consisting of 3 pieces of cheese, and the smaller one 1 piece of cheese (approx. 1 cm^3). Tray colour was selected based on dogs being able to discriminate between blue and yellow easily due to their dichromatic vision (Neitz *et al*, 1989). The colour of the tray containing each reward was randomised between subjects within the constraints of counterbalancing across the population for both colour of tray and contents. Both trays were cleaned with Safe4 disinfectant wipes between each trial.

A tape measure was attached to the floor to a length of 10m to enable the experimenter to rapidly set the trays up at the start, move the large reward tray accurately during testing, and measure the Maximum Distance Travelled (MDT) by the dog. Each tray was positioned lengthwise, one 25cm to the left, with the front of the tray parallel to the start of the tape, the other tray positioned 25cm to the right in the same way. An opaque screen (50cm high) running the length of the tape was used to ensure that once the dog had made a choice it could not change that decision and switch sides. Test subjects were positioned 1m in front of the start position of the trays.

A curtain (80cm high) was used to block each dog's visual access to the equipment when not being tested. This was positioned between the test subject and the start of the tape measure and trays. Before each trial the curtain was dropped to allow the dog to approach the test set up and after each trial the dog was taken back behind the start line and the curtain was put back up.

Dogs were restrained on a short lead that was attached to a secured, static point behind the experimental set up when not making a choice. When required to perform the task they were swapped from the short lead onto a long line and the curtain was dropped to allow the dog to approach the experimental set up and make a choice. The line was held by the researcher from behind the dog. Once a tray had been chosen, the dog was allowed to self-reward from that tray, once finished eating the lead length was reduced and the dog was then taken back behind the start line and re-secured. The curtain was then put back up and the rewards replenished and equipment repositioned (if the large reward tray had been selected). The researcher operated the curtain and managed the dog while an assistant replenished rewards and repositioned the equipment. Training and testing stages were conducted within a controlled enclosed space at the University of Lincoln Animal Behaviour and Training Centre.

4.2.3.1.3 Training protocol for spatial discounting task (SDT):

Both trays were presented simultaneously for 12 trials (dual-presentation trials). Testing only occurred if the dog met criterion by choosing the tray containing the larger reward at least 10/12 times in 2 successive sets of dual-presentation trials, the last 5 choices in each trial had to be for the larger reward tray. Dogs were given up to 5 sets of 12 dual presentation trials to reach criteria.

4.2.3.1.4 Testing: distance choice test:

Testing began following a 10 minute break after meeting criterion.

This stage involved repeated presentation of both trays (as shown in Figure 4.2). Each time the large reward tray was chosen it was moved further from the start line by 25cm. When the small reward tray was chosen, the position of the trays remained the same for the next trial; rewards were then simply replenished. This testing stage ended when the dog had chosen the small reward tray 5 times in succession. At this point, the distance to the large reward tray was measured and recorded as the MDT. All MDT measurements were taken from the front of the tray. The procedure lasted no longer than 2 hours including breaks.

All dogs were required to choose the tray containing the large reward at least once during test trials for their MDT results to be considered usable, i.e. the animal was making a rational transitive choice. Half of the population were retested within 4 weeks of the initial test (12/24 dogs).

4.2.3.1.4 Data & Statistical Analysis (SPSS 21.0):

Initial summary statistics were calculated. Linear regression analysis was then used to assess relationships between DIAS and MDT in the sample populations, and also to assess relationships between the test and retest for the OQS and MDT. Pearson's correlations were used to determine the strength of the correlation between MDT and OQS.

4.2.3.2 Simplified Field Study: SDTs with adult dogs

4.2.3.2.1 Subjects:

13 dogs aged between 2 years and 9 years were recruited from the general public to take part in the simplified spatial discounting test (SDTs), a simplified version of the SDT. For demographic information on the dogs see Figure 4.4.

Age (months)	Sex	Breed
108	M	West Highland White Terrier
108	M	Cross Breed
84	M	Cross Breed
48	F	German Shepherd
36	F	Cross Breed
48	F	German Shepherd
72	M	Border Collie
84	M	Border Collie
60	F	Jack Russel Terrier
72	F	Jack Russel terrier
96	M	Australian Shepherd
36	M	Labrador
48	M	Yorkshire Terrier

Figure 4.4 Demographic information for dogs in the Simplified Field Study Spatial Discounting Test (SDTs).

This involved the following changes:

4.2.3.2.2 Modifications for simplified field study

Modifications were made to test the protocol for field use, this is why the environment was changed, and the training simplified. Training and testing was carried out in an environment familiar to each dog, usually in the garden or driveway of their home, rather than in a controlled setting. The barrier and curtain were removed from the design of the set up. Dogs were trained to discriminate between a blue, large reward tray (3 pieces of preferred food of the subject, rather than cheese) and yellow, small reward tray (1 piece of preferred food of the subject) by presenting the large reward tray thirty times followed by the small reward tray ten times. Progression to the testing stage only occurred if dogs met criterion by choosing the large reward tray at least 8/10 times in 2 successive dual presentation trials, but for this study the last 5 choices did not need to be exclusively for the large reward.

Testing ended when the dog had chosen the small reward tray 10 times in succession rather than 5.

4.2.3.2.3 Data & Statistical Analysis (SPSS 21.0):

Linear regression analysis was used to assess relationships between DIAS and MDT. Pearson's correlations were used to determine the strength of the correlation between MDT and OQS.

4.2.3.3 Laboratory Study 2: SDT with Young Dogs

4.2.3.3.1 Subjects:

25 dogs aged between 2 and 9 months were recruited from the general public, 24 of which reached criterion for data inclusion. Half of the subjects (12) were retested 4-6 weeks after initial testing. For demographic information on the dogs in the study see Figure 4.5.

Age (months)	Sex	Breed
3	F	German Shepherd
2	F	Labrador
2	F	Golden Retriever
3	F	Cavapoo
5	F	Labrador
4	M	Miniature Schnauzer
4	M	Labrador
3	F	Border Terrier
4	M	Springador
3	M	Shih Tzu
4	F	Chinese Crested Powder Puff
3	F	American Bulldog
2	M	Springer Spaniel
3	F	Crossbreed
3	F	Lurcher
3	M	Cockapoo
3	M	Springer Spaniel
2	M	Large Munster lander
4	F	Malinois
4	F	Greyhound
4	M	Crossbreed
3	M	Cockapoo
4	M	Border Collie

Figure 4.5. Demographic Information on the Young Dogs in the Spatial Discounting Test.

4.2.3.3.2 Modifications for study with young dogs

The SDT protocol from experiment 1 was followed, except that the pieces of cheese were $0.5cm^3$, i.e. $1/8^{th}$ of the volume to account for the dogs being younger and smaller than the adults.

4.2.3.3.3 Data & Statistical Analysis (SPSS 21.0):

Linear regression analysis was performed on the data to assess relationships between DIAS and MDT in the sample populations, and also to assess relationships between the test and re test for the OQS and MDT. Pearson's correlations were used to determine the strength of the correlation between MDT and OQS.

4.3 Results

4.3.1 Laboratory Study 1: SDT with adult dogs

Of the 25 dogs recruited, 24 (96%) reached criterion and completed the SDT. 1 adult dog was subsequently excluded from analysis due to being diagnosed with hip dysplasia shortly after the testing. Accordingly 23 out of the 24 dogs provided data for analysis of convergent validity with DIAS. Eleven out of 12 dogs provided data for test-retest analysis.

The mean MDT for adult dogs was 3.27m +/- 1.74m and the mean OQS for adult dogs was 0.55 +/- 0.1. A significant negative correlation ($r = -0.46$, $p = 0.028$, adj $R^2 = 0.171$) was found between OQS and MDT. See Figure 4.6.

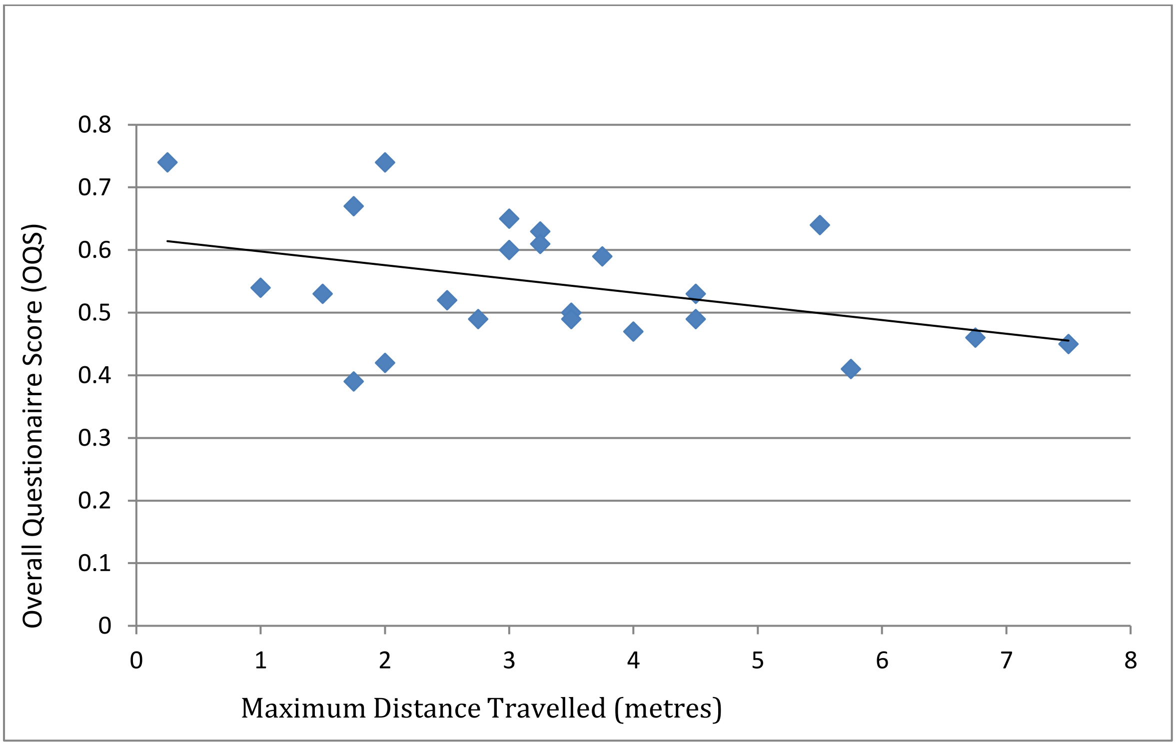


Figure 4.6. Relationship between Overall Questionairre Score (OQS) and Maximum Distance Travelled (MDT) in adult dogs in the Spatial Discounting Task.

Test Retest:

The mean OQS at test was 0.55 and the mean OQS at retest was 0.53. The relationship between OQS at test and retest in adult dogs was found to be significant (r = 0.949, p<0.0001, adj. R^2 = 0.89).

The relationship between MDT at test and retest in adult dogs was also significant (r = 0.859, p=0.001, adj. R^2 = 0.71). The mean MDT at initial testing was 3.27m and the mean MDT at retest was 3.55m (Figure 4.7).

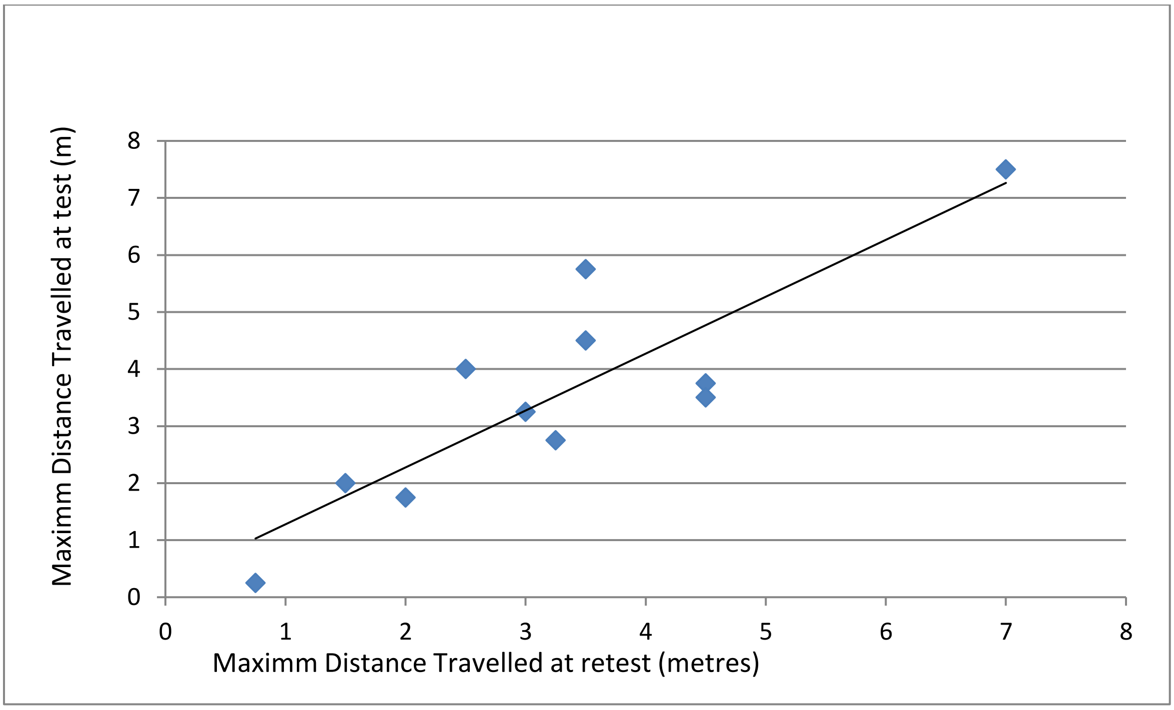


Figure 4.7 Relationship between Maximm Distance Travelled at test and retest in adult dogs in the Spatial Discounting Test.

DIAS Factor Analysis:

A significant relationship was also found between MDT and DIAS Factor 1 (r = -0.463, p=0.026, adj. R^2 = 0.177 (Fig 4.8).

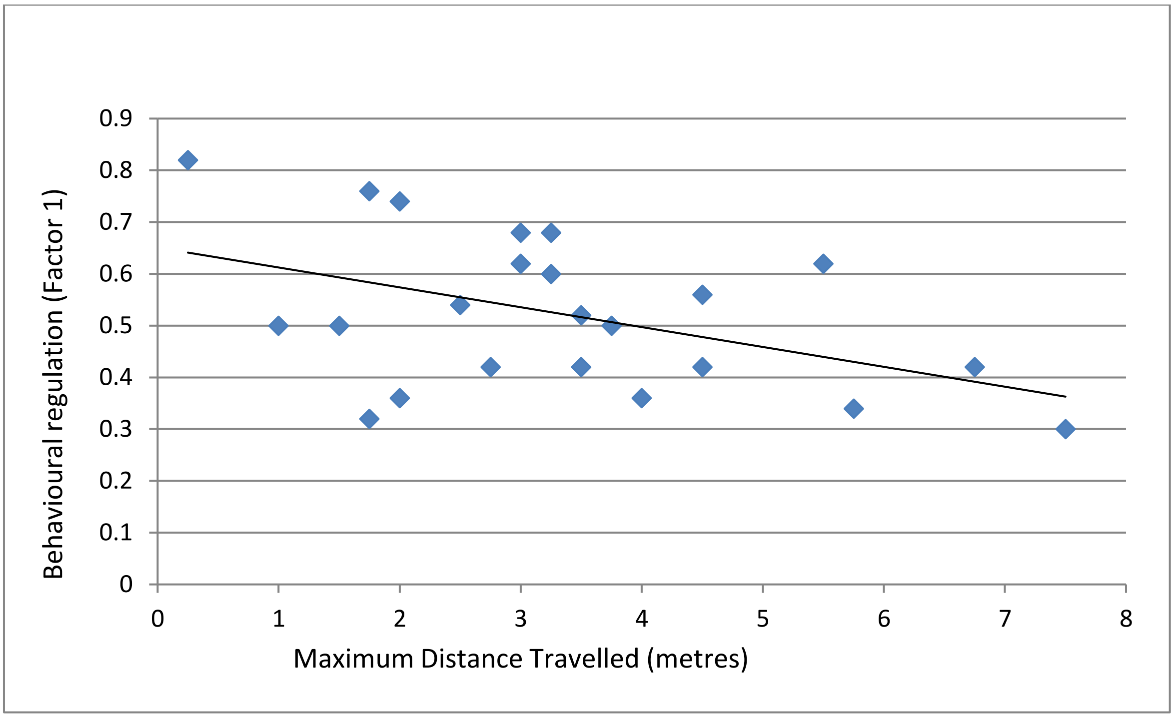


Figure 4.8. Relationship between Maximum Distance Travelled and Behavioural Regulation (Dog Impulsivity Assessment Scale Factor 1) in adult dogs in the Spatial Discounting Test.

No Significant relationships were found between MDT and DIAS Factor 2 (p=0.971) or MDT and Factor 3 (p=0.296).

4.3.2 Simplified Field Study: SDTs

All 13 dogs recruited met criteria for testing and all provided data for analysis. The mean MDT was 1.88m +/- 1.64m and the mean OQS was 0.53 +/-0.07. A significant negative correlation ($r=-0.61$, $p=0.028$, adj. $R^2 = 0.313$) was found between OQS and MDT (Figure 4.9).

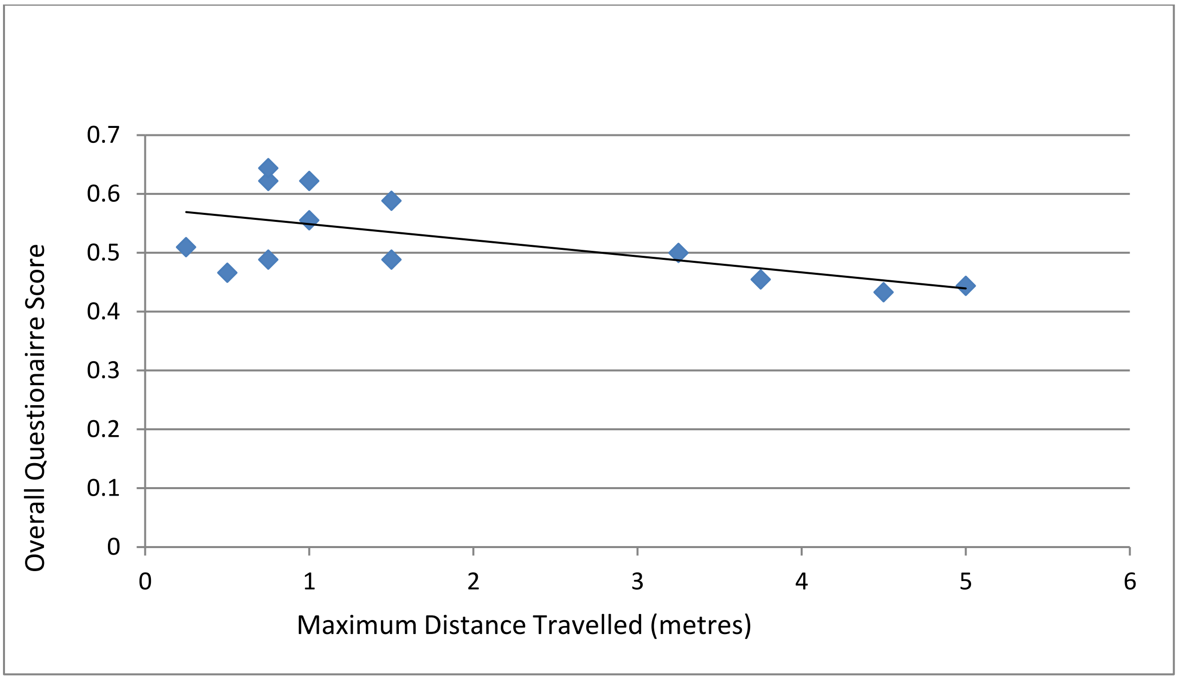


Figure 4.9. Relationship between Overall Questionairre Score (OQS) and Maximum Distance Travelled (MDT) in adult dogs in the Simplified Spatial DIscoutning Test (SDTs).

No Significant relationships were found between MDT and Factor 1(p=0.162), Factor 2 (p=-0.192) or Factor 3 (p=0.154).

4.3.3 Laboratory Study 2: SDT with young dogs

Of the 25 young dogs recruited, 24 (96%) reached criterion and completed the behaviour test (SDT). Twenty three out of the 24 young dog data were used for analysis, 11/12 retest data were analysed. One of the young dogs was excluded from analysis after having completed the testing as they had come into season during the retest and this may have impacted on their results.

The mean MDT in young dogs was 2.70m +/-1.94m and the mean OQS was 0.54 +/-0.06. The relationship between OQS and MDT for puppies was not significant ($r = -0.053$, $p=0.811$, adj. $R^2 = -0.062$ (Figure. 4.10).

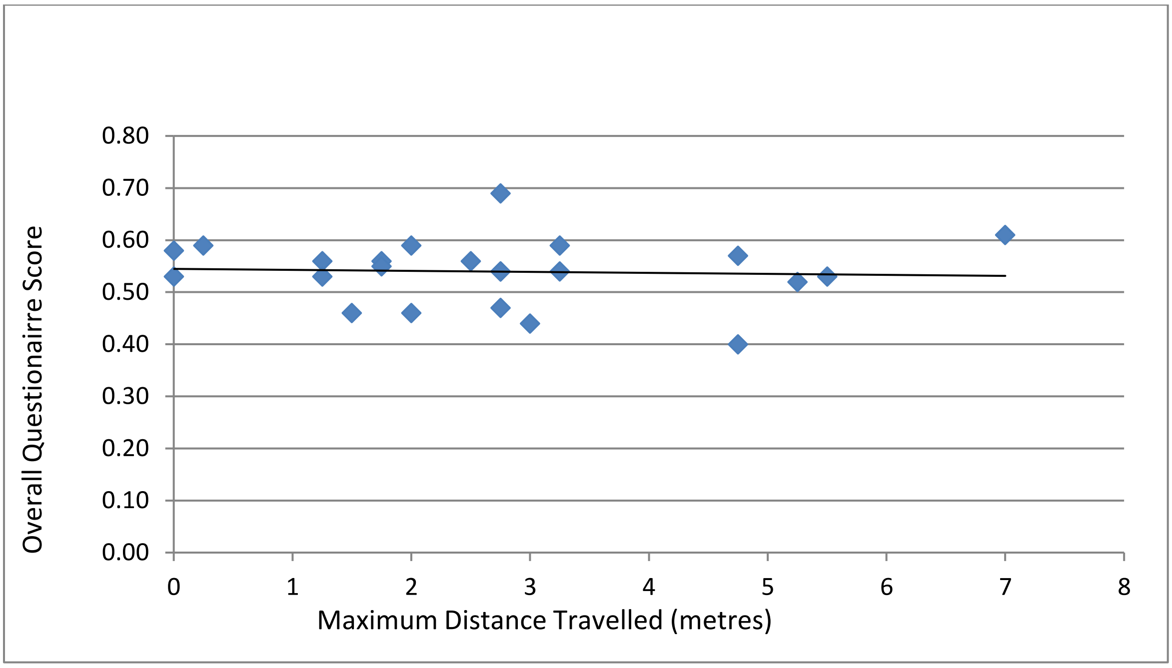


Figure 4.10. Relationship between DIAS Overall Questionairre Score (OQS) and Maximum Distance Travelled (MDT) in young dogs in the Spatial Discounting Task.

Test Re test:

The relationship between OQS at test and retest in young dogs was significant ($r = 0.673$, $p=0.023$, adj. $R^2 = 0.393$). The mean OQS at test was 0.54 and the mean OQS at retest was 0.57. (Figure 4.11).

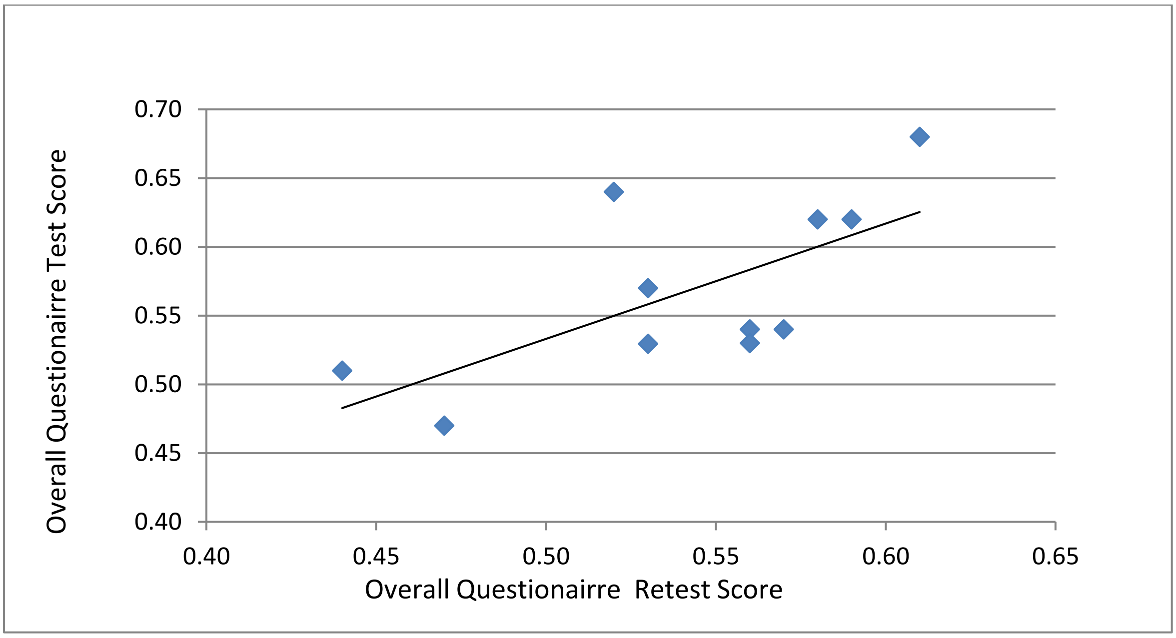


Figure 4.11. Relationship between DIAS Overall Questionairre Score at test and retest in young dogs in the Spatial Discounting Task.

The relationship between MDT at test and retest in puppies was not significant (r = 0.414, p=0.15, adj. R^2 = 0.115). The mean MDT at test was 3.09m and the mean MDT at retest was 2.80m (Figure 4.12).

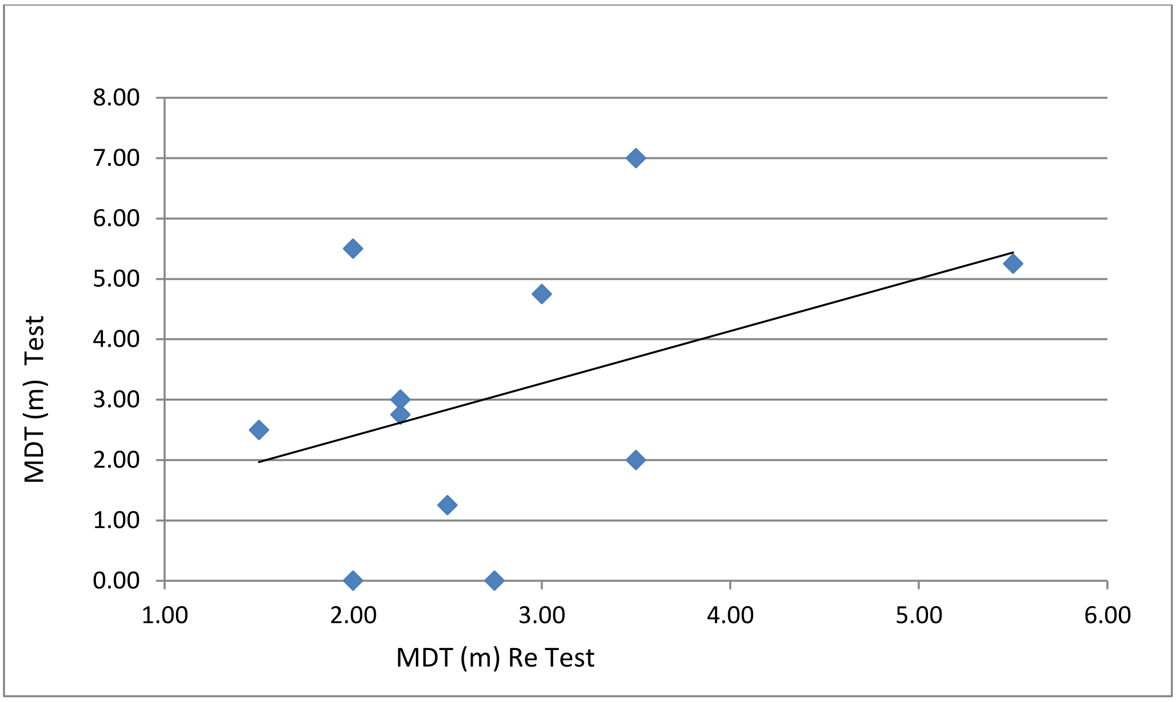


Figure 4.12. Relationship between Maximum Distance Travelled (MDT) at test and retest in young dogs in the Spatial Discoutning Test

A significant positive correlation (r=0.62, p=0.044, adj. R^2 = 310) was found between DIAS Factor 3 and MDT in young dogs in the retest (Figure 4. 13), but not for the initial test (r = 0.421, p=0.197, adj. R^2= 0.086.

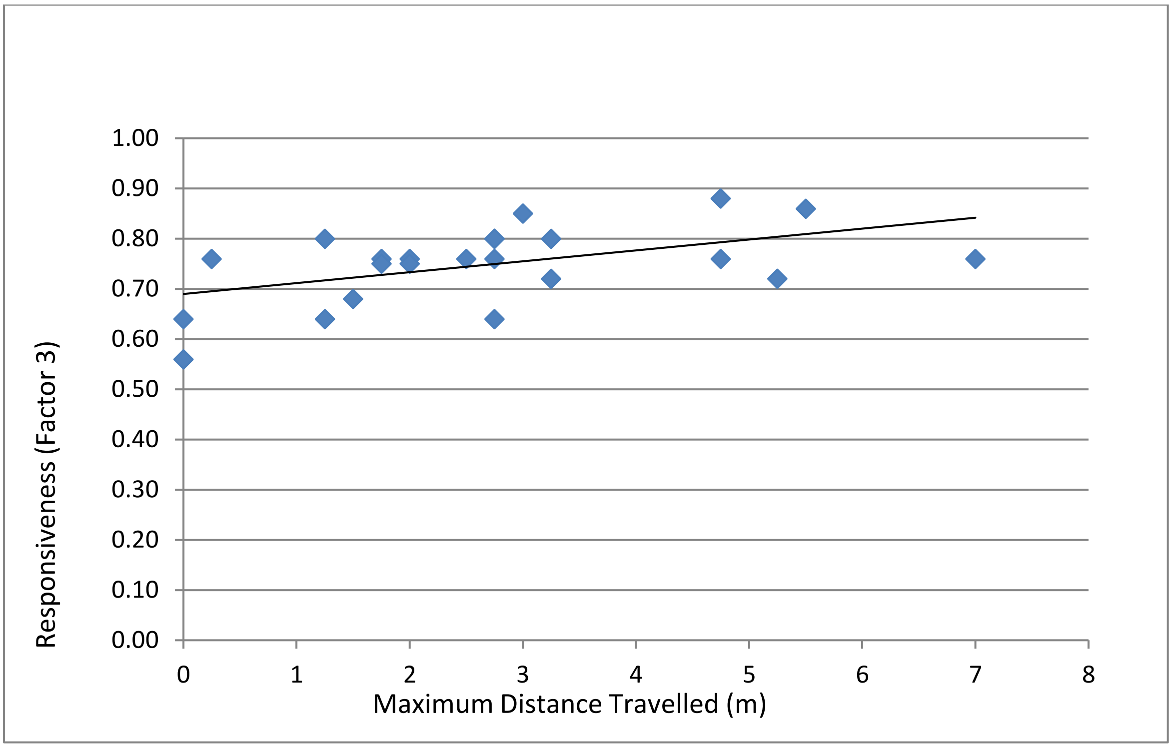


Figure 4.13. Relationship between Maximum Distance Travelled and DIAS Factor 3 Responsiveness in Young Dogs in the Spatial Discounting Test.

No significant relationships were found between MDT and DIAS Factor 1 (p=0.59) or Factor 2 (p=0.84).

4.4 Discussion

This study met its aims in terms of developing a spatial discounting task suitable for assessing impulsivity in dogs over 2 years of age. The spatial discounting task was validated against DIAS in adult dogs, with spatial discounting task and DIAS reliable for assessing impulsivity in adult dogs. There are still challenges relating to the assessment of impulsivity in young dogs, while the DIAS was found to be consistent, the age at which the spatial discounting task becomes consistent and so can be used to assess trait impulsivity in dogs is unknown.

Convergent validity was found between the behaviour test and OQS (p=0.028) in the controlled laboratory study, as well as the behaviour test and OQS Factor 1 (behavioural regulation) (p=0.026). In practical terms this meant that dogs that travelled a shorter distance before switching had higher impulsivity scores compared to those that travelled further in the behaviour test; this effect has previously been demonstrated in a temporal discounting task (Wright *et al*, 2012). The spatial discounting task also showed consistency with MDT (p=0.001) measures used in the controlled laboratory test were stable over after 4-6 weeks for adult dogs. In light of this, it is suggested that the behaviour test could be used to predict impulsivity as measured by OQS in adult dogs when it is not suitable for the DIAS to be completed, for example when a dog does not have a consistent caregiver to answer the questionnaire. Although the results were not directly compared, the test yields similar results to previous research which used a temporal discounting paradigm (Wright et al, 2012) to assess impulsivity.

Spatial delay to reward allows the animal to visualise the comparable distances to the two options, unlike what happens in the temporal delay. This would suggest that the way the delay is represented between the two methods is different. Nonetheless a similar pattern was observed in terms if delay tolerance and impulsivity in the spatial and temporal task, although this does not equate to convergence. Patience, the tolerance of long temporal delays for gain has been suggested to be a largely human trait, with reward value shown to influence the temporal delay tolerated (Rosati *et al*, 2007). This could be investigated using the spatial discounting task in dogs to see how value influences the spatially induced delay they tolerate. Reward ratio has been shown to be more important than absolute values, with a 1:3 ratio shown to be an effective difference in primates (Long & Platt, 2005). This method was also used in a temporal delay task with dogs (Wright *et al*, 2012). Using absolute reward value or type of reward rather than ratio to examine discounting could also further our understanding of how dogs perceive different rewards. Manipulating the reward differential may alter the maximum delay they tolerate, and so used to shorten the behaviour test, perhaps providing an even more concise way to assess impulsivity.

Previous studies in a range of species have compared spatial and temporal discounting-but not in terms of incremental delay tolerance (Stevens *et al*, 2005, Long & Platt, 2005). Differences have been found between spatial and temporal delay tolerance within and between primate species, but have not examined individual delay tolerance based on temperament profiling or links between spatial discounting and temperament in non-human animals (Stevens *et al*, 2005). Relationships between temporal discounting and temperament traits relating to impulsivity have been observed in a variety of species (Ostaszewski, 1996. Perry *et al*, 2007. Ainslie, 1974. Wolff *et al*, 2002. Tobin et al, 1996. Evans & Beran, 2007. Rosati *et al*, 2007. Bickel et al, 2007). Our research shows individual differences in tolerance to spatial delay of reward in dogs that appears to be correlated with how impulsive the individual is, and previously a similar relationship has been found for temporal delay in dogs (Wright *et al*, 2012). These results show that the spatial discounting task is a simple tool that can be used to assess impulsivity in adult dogs. The spatial discounting task is a faster and easier method for assessing impulsivity in dogs compared to previous laboratory assessments using temporal discounting. The maximum time for completion of the task was one 2 hour session and at least 96% of dogs recruited across the 3 studies were able to complete testing. Being quick and simple for the dogs to complete the task, extends the way the test can be utilised. Results of the field adaptation also show that a simplified version of the spatial discounting task can be used in a less controlled setting to assess impulsivity. The simplified field study showed good convergent validity with OQS ($p=0.028$), making it more practical to assess impulsivity in dogs in their own environment. No relationship was found between the field behaviour test and Factor 1 (Behavioural Regulation), while this could be a result of the smaller sample size, it could also indicate that the field test does not provide the same level of sensitivity relating to assessment of the trait's components compared to the laboratory test.

High impulsivity has been linked to behaviour problems resulting from a lack of self-control or inability to tolerate frustration. This may be particularly pertinent to working dogs and their selection but also to examining the extent to which self-control training regimes can influence impulsivity. While traits are relatively consistent over time and across context, it is not known how much they can be manipulated, if at all, by training. Previous research into this has had limitations, since it has been shown that training history impacts on ability to learn the task and so results in a self-selection bias in the population completing the test (Wright *et al*, 2012). By contrast the spatial discounting task required little to no training of the dogs. The simplicity and robustness of the task (shown here by the reliability of the simplified field adaptation) enables it to be used more easily prior to and post interventions aimed at improving self-control. This has important implications and could enable us to shape individuals towards a more desirable temperament. Following the assumption that impulsivity is in some way linked to rejection or withdrawal, this could make this task a useful tool for use in assessing working dogs.

Fast and simple assessment also enables the spatial discounting task to be used more widely than previous tests. This makes it a practical option for the selection of working dogs. Identification of how trait impulsivity presents in successful or unsuccessful working dogs could make this test useful in the early identification of risk of behaviour problems that may present in the field that could be missed during conventional selection and training processes. Although as it is an assumption that impulsivity links to working dog performance further investigation into the trait in working populations would be necessary before applications could be applied.

There are still challenges to assessing impulsivity in young dogs, the age at which the trait becomes stable and measurable remains unknown. No significant relationship was found between young dog MDT and OQS ($p=0.811$) in laboratory study 2, although the OQS ($p=0.023$) was stable over time (4-6 weeks later) at this age, the MDT ($p=0.15$) was not. This suggests that the age at which the behaviour test can be used to predict the OQS score remains unknown. This could be because the trait itself is not stable at this age, but owners' opinions may become fixed early on. However, as the behaviour test appears to be able to predict impulsivity score in adult dogs, then logically there will be a point at which the test becomes predictive, and there are potentially other factors influencing the behaviour test in young dogs. There is the potential that the young dog population was unable to perform to the same extent as the adults in the behaviour test. It could also be suggested that the behavioural expression of impulsivity is not as fixed in young dogs as it is in adult dogs meaning that there is more room to shape the expression of the trait in early months. Due to the varied ages of the young dogs sampled in this study there is insufficient data to investigate a specific age effect. More research is needed to establish if a test can be devised that can predict overall questionnaire score for impulsivity in young dogs as well as determine at what age the spatial discounting task becomes predictive.

A longitudinal study assessing young dogs at regular intervals into adulthood would be useful to establish an age range within which the trait stabilises. This could be achieved through distributing the DIAS to puppy owners and tracking their puppies on a monthly basis to identify the point at which the results of the questionnaire remain the same for a set period of time. However, it would also be necessary to examine the trait empirically such as through the spatial discounting task in a longitudinal study to determine at what age the maximum distance travelled (or other measure of delay tolerance) correlates with their overall questionnaire score for impulsivity, and at what age the MDT in puppies is temporally stable. However there could be an effect of learning if the test was repeated on multiple occasions. It would also be potentially useful to examine urinary metabolites of serotonin and dopamine in young dogs being tested to look for further correlations with either factors within the DIAS or maximum distance travelled in the spatial discounting task, as these metabolites were found to correlate with both the maximum delay tolerated in the operant behaviour test for adult dogs, as well as their DIAS scores (Wright *et al*, 2012).

4.5 Conclusions:

The spatial discounting task has been shown to be a valid measure of OQS for impulsivity based on the DIAS in adult dogs. It is a preferable test to use compared to previous behavioural methods of assessing trait impulsivity (Wright *et al*, 2012) as the spatial discounting task takes less time to train and test the dogs and results in a very high completion rate. The simplified field version of the spatial discounting task can be used to more roughly assess impulsivity in dogs where there are limitations on experimental set up, this enables the tool to be more widely used in a field setting. This makes the test more practical for use in assessing impulsivity in dogs where the DIAS cannot be completed and could be useful for objectively selecting dogs for working roles. While the OQS was found to be a reliable tool for measuring impulsivity in young dogs of known history, more research is needed to establish the age at which the spatial discounting task shows convergent validity with OQS score for impulsivity to allow identification of an age at which we can reliably test impulsivity using the behaviour test.

Chapter 5: What distinguishes a successful working dog from an unsuccessful one?

Synopsis

The previous chapter developed simple and effective measurement methods for behaviourally assessing impulsivity in dogs. Theoretically impulsivity, as well as other temperament traits linked to resilience and distractibility such as core affect, will influence success in work and could provide a useful contribution to working dog selection. Previous research has focused on identifying factors related to passing certification or completing training, but has not examined factors relating to dogs being withdrawn from service or working successfully for the intended duration of their working life. This is likely contributing to working dog withdrawal as current selection processes still have room for improvement. Investigating the characteristics of successful vs unsuccessful working dogs should help us to fill some of these gaps in knowledge. This chapter focuses on two areas for which we have validated tools for assessing aspects of temperament; impulsivity and core affect, and seeks to identify possible differences in these traits in working dogs that are successful compared with those that are withdrawn from service for behavioural reasons. Surveys were distributed to the working dog sector to gather demographic information about the dogs from their handlers, and also to get handlers to complete the DIAS for impulsivity and the PANAS for core affect for their dogs. Scores for the elements of the PANAS and DIAS were then calculated and compared between active working dogs and withdrawn working dogs from a dataset of UK police dogs. Further comparisons were made between the UK police dog population and the UK military dog population, additional comparisons were made to the pet dog population based on an earlier dataset (Wright *et* al, 2012). Key differences were found in the element 'Responsiveness' from the DIAS scale and 'Energy & Interest' from the PANAS scale between UK active police dogs and UK withdrawn police dogs from this dataset, with the withdrawn dogs scoring lower for both components. The pet dog population also significantly differed to the UK active police dog sample with significantly lower scores for the elements 'Behavioural Regulation' and 'Responsiveness' within impulsivity, and scoring lower for 'Energy & Interest' within core affect. The pet dog population also scored higher than the UK active police dog sample in terms of 'Negative Activation' of core affect from the PANAS. Active police and military working dogs were found to have statistically similar profiles in terms of all the elements of PANAS and DIAS, although insufficient access to withdrawn military working dogs prevented further comparisons. This suggests that there is the potential for positive selection for certain qualities in working dogs and these traits might be different in dogs withdrawn from service. Although the causal nature of the relationship cannot be determined by such methods, this work provides insights for future testable hypotheses to move the field forward.

5.1 Introduction:

5.1.1 Temperament

Temperament, defined in Chapter 1 as the individual differences in behavioural responses, relatively consistent across time and context and grounded in emotional states, core affect and related behavioural processes, consists of adaptive traits, with different combinations being more advantageous in different environments. Traits exist along a spectrum in nature since no single level is consistently the best in all situations (Sheppard & Mills, 2003). To this end temperament traits have been studied to understand the impact they have on fitness in wild populations and the role natural selection has on trait development and persistence of individual differences in natural populations (Reale *et al*, 2007). Certain temperament traits appear to be present and have persisted in a number of species through evolution (Reale *et al*, 2007), which suggests that they are linked to success in terms of survival and fitness (Dingemase & Reale, 2005). A number of temperament and personality traits have been studied in order to try to identify those associated with success (Slabbert & Odendaal, 1999. Mann, 1959. Seibert *et al*, 1999. Coleman *et al*, 2005. Sinn *et al*, 2010). In particular, desire for work and distractibility have been identified as important in terms of scent detection dog certification in Japan, with dogs possessing a stronger desire for work and low levels of distractibility being more likely to pass certification (Maejima *et al*, 2007). Both of these attributes can be traced back to elements of temperament relating to core affect and impulsivity as discussed in Chapter 1.

5.1.2 Temperament, Personality & Performance

Aspects of individual differences including personality and temperament have been linked to career success (Seibert, 1999. Thompson, 2005) with specific traits predictive of academic success in humans (Colom *et al*, 2007. Komarraju *et al*, 2009, Chamorro-Premuzik & Furnham, 2003. Palisin, 1986) while in animals such traits have been linked to learning performance factors such as training success in primates, with individuals categorised as “exploratory” learning a task faster compared to those categorised as “inhibitory” in terms of temperament as defined within their research (Coleman *et al*, 2005). Understanding this relationship has enabled researchers to select animals suitable for training and research projects, and to identify individuals in need of alternative training methods to optimise performance (Coleman *et al*, 2005).

Traits associated with success could be used to predict performance outcomes in a variety of contexts. Relationships between temperament and performance have been documented across a number of disciplines ranging from psychology to sports science (Judge & Ilies, 2002. Komarraju *et al*, 2009. Barrick *et al*, 2005. Roberts *et al*, 2007. Morris, 2000) and across species from humans (Seibert & Kraimer, 2001), to cats (McCune, 1995.), donkeys (French, 1993.) and dogs (Wright *et al*, 2012). Developing an understanding of temperament traits that are related to performance or success has the potential to allow prediction of important outcomes and ability to succeed in certain environments or tasks, as well as understand the impact temperament has on behaviour.

Aspects of temperament have been found to be related to personality disorders in humans (Fiedler *et al*, 2004) behavioural problems in humans (Sanson *et al*, 2004) and dogs (Wright et al 2011). Such behaviour problems sometimes arise from a general inability to cope well with certain situations or stimuli. Temperament in animals has been linked to coping ability with individuals possessing certain traits showing greater ability to adapt to environmental challenges such as adverse weather conditions. Certain traits have also been linked to specific coping styles in animals (Seaman *et al*, 2002) and in humans to predict behavioural adjustment throughout the early years as well as the ability of humans to cope with (Palisin, 1986) and adapt to (Fiedler *et al*, 2004) change. Ability to cope will of course also impact on performance especially under challenging conditions. Behaviour tests were discussed extensively within Chapter 3, but questionnaires can also be useful when it comes to assessing traits of relevance to working dogs.

5.1.3 Questionnaire assessment of behaviour

Behavioural assessment for working dogs are not limited to behavioural tests. Dog behaviour questionnaires usually ask questions about behaviour across a number of contexts, and use this information and various scoring methods to identify behavioural tendencies. Some questionnaires focus on a wide range of traits, while many focus on factors such as fear, aggression or anxiety (Temesi *et al*, 2014). However, behaviour tests are often the preferred method as with using a trained observer to assess behaviour they are likely to have a better understanding of normal and abnormal dog behaviour in testing situations, due to the volume of dogs they assess. Questionnaires can also be doubted when they rely on a consistent caregiver to complete them, as prior knowledge and relationship with the dog could bias some answers (Weiner & Haskell, 2016). Where possible, questionnaires can be used alongside behaviour tests to look for correlates, and this is thought to improve outcomes (Weiner & Haskell, 2016). A number of questionnaires have been used in this way (Weiner & Haskell, 2016), for example The Dog Personality Questionnaire (Mirko *et al*, 2012) pinpointed 4 personality traits through questionnaire measured and was also examined using behaviour testing although only 3 of the traits correlated to test behaviour (Mirko *et al*, 2013). The Monash Canine Personality Questionnaire (MCPQ) and its revised version (MCPQR), is an owner rating questionnaire for personality in dogs (Ley *et al*, 2008). This questionnaire has been examined in pet dogs and has been shown to demonstrate correlations between personality factors defined using the questionnaire and behaviour in a dog park (Carrier *et al*, 2013).

The Canine Behavioural Assessment and Research Questionnaire (C-BARQ) is a commonly used questionnaire for assessing dog behaviour using a 5 point scale. Studies have found correlations between C-BARQ scores and behavioural test results (for aggression-related traits, van den Berg *et al*, 2006; and sociability-related traits, De Meester *et al*, 2008). Studies have also linked C-BARQ scores to success as guide dogs or service dogs (Duffy & Serpell, 2012; Foyer *et al*, 2014) as well as rescue dog outcomes (Duffy *et al*, 2014). However, the predictive ability of the C-BARQ for military working dog selection based on early life C-BARQ traits is not exceptionally high; for example, in a study of German Shepherd military dogs the highest correlation between early life C-BARQ traits was 0.36 (Foyer *et al*, 2014).

The PANAS (Positive and Negative Activation Scale (Shepherd and Mills 2002) is a tool used to assess sensitivity to rewards (positive distractions - temptations) and aversives (negative distractions – anxieties). The PANAS is broken down into measuring elements of "Negative Activation", a standalone element, and "Positive Activation" which is made up of three factors, "Energy & Interest" "Persistence", and "Excitement". The PANAS has been found to be robust in animals as young as 10 weeks of age (Sheppard and Mills 2002) and has predictive validity in relation to problems behaviours associated with fear and distractibility (Mills *et al*., unpublished data). The importance of fear and distractibility in working dog failure is widely acknowledged (see Chapter 1), and the PANAS could be a useful tool to assist with measuring these in working dogs.

The DIAS (Dog Impulsivity Assessment Scale (Wright *et al*, 2011)) is a tool for assessing impulsivity in dogs. It measures impulsivity firstly as an overall score for the trait, but also comprises of three factors. Factor 1 Behavioural Regulation, Factor 2 Aggression and Response to Novelty, and Factor 3 Behavioural Responsiveness. The DIAS has been validated with both behavioural (delayed reward choice task) and physiological (urinary metabolites of serotonin and dopamine) measures in dogs (Wright *et al*, 2012). Additionally, when assessed in this way, it has been shown to be stable for up to 6 years (Reimer et al, 2014). As discussed in Chapter 1, Impulsivity is potentially relevant to the assessment of distractibility (Peremans *et al*, 2002), and is expressed in a range of behaviours linked with inhibitory control, trainability and focus (Wright *et al*, 2012), which also relate to working dog performance as outlined in Chapter 1.

The PANAS and the DIAS are tools which can be used to measure traits of relevance to working dogs. While these have been utilised in pet dogs, applications to working dog populations has never been attempted, investigating these tools for use in working dogs could provide additional measures to assist current selection procedures.

5.2 Aims

The aim of this study is to identify if there is a relationship between any of the components of core affect (based on the PANAS) or impulsivity (based on the DIAS) and working dog success versus withdrawal from the field.

5.3 Methods

5.3.1 Survey

A secure online survey was created for PANAS and DIAS questionnaires along with dog demographic information including whether or not the dogs have been withdrawn from service due to behavioural concerns. Handlers were asked to score their dog's performance in training and again in the field, using a Likert scale ranging from 1 (poor) to 5 (excellent). The survey was available online at http://www.uoldogtemperament.co.uk/workingdogs/ between March 2014 and October 2015 (See Appendix 3 for paper version). The link and password were distributed to dog handlers through contacts at Greater Manchester Police dog unit, British Transport Police, and via Dstl to the military, and to the Dutch working dog sector via the Netherlands specialist canine unit.

5.3.4 Hypotheses

It was hypothesised that there would be significant differences in terms of scores on the DIAS for impulsivity and the PANAS in terms of positive and negative activation and/or their component factors between dogs currently in active service and dogs that had been withdrawn from service for behavioural reasons. More specifically it is hypothesised that dogs scoring higher for Negative Activation scores from the PANAS and for Aggression and Response to Novelty from the DIAS will be linked to withdrawal from service, based on the literature surrounding fear and failure in working dogs as both of these tap into a fearful response.

As working dogs are selected for certain traits it is hypothesised that working dogs will differ in PANAS and DIAS scores compared to dogs from the general population.

5.3.5 Data analysis

Statistical analysis was performed using SPSS 21.0.

KS tests for normality showed the data to be parametric, and paired t tests were used to compare training and field score in active dogs, training and field score in withdrawn dogs, and training and field score in the entire population. These scores were then compared between active and withdrawn dogs using independent samples t tests, and also looked at to identify any correlations between handler score and the questionnaire scores from the PANAS and DIAS using Pearsons Correlations.

All data were analysed to compare active and withdrawn dogs in terms of their temperament profiles generated from the PANAS and DIAS questionnaires. Kolmogorov-Smirnoff tests for normality were used as well as Levene's test of equal variance. As the data were found to be parametric, unrelated, between subjects t-tests were then used to compare the various populations of withdrawn versus active dogs and comparisons between working dogs and the pet dog population.

Active dogs from the military dog sample and the Netherland dog sample were analysed against the UK active police dog sample to look for differences using independent samples tests, and correlations to look for any relationships between the PANAS and DIAS questionnaire scores for the data sets.

A sub population was taken from the police dog study sample to match individuals for breed, age, sex, neuter status and working role as potential confounding factors. This matched pairs sample was then assessed using independent sample tests to look for differences in terms of the components of temperament from the PANAS and DIAS questionnaires.

Due to the small sample sizes of working dogs, effect sizes (Hedges G) were calculated for comparisons between these populations to establish the size of effects and evaluate the risk of type 2 statistical error with non-significant results.

5.4 Results:

5.4.1 Samples

Eighty-seven UK police dog handlers responded to the survey. After excluding UK police dogs who had a medical condition (n = 2; withdrawn) and those with incomplete questionnaires (active: n = 5; withdrawn: n = 1) data were collected and analysed from 79 dogs (active: n = 63; withdrawn: n = 16). Of these dogs 63 were currently in active service (*Breed:* Working collie: 5 (7.9%), German shepherd: 31 (49.2%), Labrador retriever: 6 (9.5%), English springer spaniel: 15 (23.8%), Belgian Malinois: 2 (3.2%), Cross breed: 4 (6.3%); *Sex:* Male entire: 33 (52.4%), Male neutered: 16 (25.4%), Female entire: 7 (11.1%), Female neutered: 7 (11.1%); *Age:* 4.81 years ± 2.34; Mean ± SD) and 16 were withdrawn from service (*Breed:* German shepherd: 6 (37.5%), Labrador retriever: 1 (6.3%), English springer spaniel: 5 (31.3%), Belgian Malinois: 1 (6.3%), Cross breed: 3 (18.8%); *Sex:* Male entire: 8 (50%), Male neutered: 3 (18.8%), Female entire: 3 (18.8%), Female neutered: 2 (12.5%); *Age:* 4.53 years ± 2.53).

Twenty military dog handlers responded to the survey. All of the completed questionnaires for military dogs made reference to animals in active work, no exclusions were made (*Breed:* German shepherd: 6 (30%), Labrador retriever: 2 (10%), English springer spaniel 1 (5%), Belgian Malinois: 9

(45%), Cross breed: 2 (10%); *Sex:* Male entire: 11 (55%), Male neutered: 2 (10%), Female entire: 3 (15%), Female neutered: 4 (20%); *Age:* 5.00 years ± 1.97).

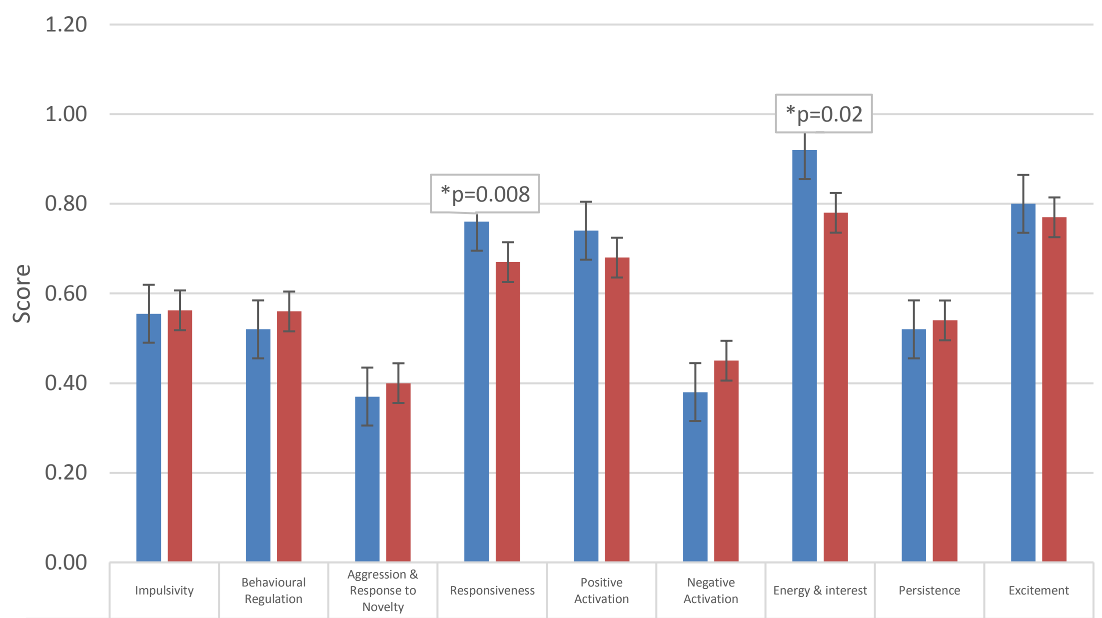
Twenty-eight police dog handlers in the Netherlands responded to the survey, no exclusions were made, all of the completed questionnaires made reference to dogs in active work (*Breed:* Dutch herder: 28 (100%); *Sex:* Male entire: 11 (55%), Male neutered: 2 (10%), Female entire: 3 (15%), Female neutered: 4 (20%); *Age:* 4.32 years ± 1.92).

Comparisons were also made with previous research in the pet dog population. For the DIAS, this consisted of 560 pet dogs reported on by Wright et al. 2011: Age range: 3 months-16.5 years. Sex: male neutered *n* = 199, (34.9%), female neutered *n* = 187, (32.7%), male entire *n* = 103, (18.0%), female entire *n* = 82, (14.4%), pedigrees *n* = 435, (76.2%) from 107 different breeds; the remainder *n* = 136, (23.8%) were crossbreeds. For the PANAS this consisted of 343 pet dogs reported on by Sheppard & Mills, 2002: Age range, 3 months – 17 years. Sex: male neutered *n* = 71 (20.7%), male entire *n* = 113 (32.94%), female neutered *n* = 89 (25.95%), female entire *n* = 67 (19.53%). The data from these studies (Wright et al. 2011, Sheppard & Mills, 2002) were collected using online and postal survey completion methods, and recruited through a range of sources including university press, dog shows, retail outlets and dog training clinics.

5.4.2 Police dog sample:

A significant difference was found between the DIAS score for "Responsiveness" in active and withdrawn working dogs in the study population. See Figure 5.1 below. Active dogs (0.76 +/- 0.1) scored significantly higher than withdrawn (0.67+/-0.14) dogs (t=2.737, p=0.008). No significant differences were found between active and withdrawn dogs in terms of the other DIAS factors, OQS (t=-0.281, p=0.78), Behavioural Regulation (t=-0.597, p=0.34) or Aggression and Response to Novelty (t=-0.956, p=0.34).

Active dogs also scored significantly higher (0.92 +/- 0.09) on the PANAS Positive Activation sub factor of "Energy & Interest" compared to withdrawn dogs (0.78 +/- 0.22) (t=2.567, p=0.02). No significant differences were found between any of the other PANAS factors either, Positive Activation (t=1.343, p=0.197), Negative Activation (t=-1.378, p=0.185), Persistence (t=-0.277, p=0.782) or Excitement (t=0.561, p=0.581). For mean scores see Figure 5.1.



	Impulsivity	Behavioural Regulation	Aggression & Response to Novelty	Responsiveness	Positive Activation	Negative Activation	Energy & interest	Persistence	Excitement
Active	0.55	0.52	0.37	0.76	0.74	0.38	0.92	0.52	0.80
Withdrawn	0.56	0.56	0.40	0.67	0.68	0.45	0.78	0.54	0.77

Figure 5.1. A comparison of the temperament components of impulsivity and core affect between active and withdrawn UK police dogs.

5.4.3 Matched pair's analysis:

A sample of 22 dogs (11 active and 11 withdrawn) from the UK police dataset were matched on 5 factors (age +/- 1 year, sex, breed, neuter status, working role).

A significant difference was found between the DIAS score for "Responsiveness" in active (0.77 +/- 0.09) and withdrawn (0.65 +/- 0.14) working dogs using independent t tests (t=-2.475, p=0.037). No significant differences were found between active and withdrawn dogs in terms of the other DIAS factors, OQS (t=0.940, p=0.369), Behavioural Regulation (t=1.075, p=0.307) or Aggression and Response to Novelty (t=0.165, p=0.873).

Active dogs also scored significantly higher on the PANAS factor of "Energy & Interest" compared to withdrawn dogs based on independent t test analysis (0.9 +/- 0.08 for active dogs compared to 0.73 +/- 0.23 for withdrawn, t= -2.24, p=0.049). No significant differences were found between any of the other PANAS factors either, Positive Activation (t=-1.43, p=0.181), Negative Activation (t=0.85, p=0.407), Persistence (t=0.77, p=0.459) or Excitement (t=-1.70, p=0.119). See figure 5.2 below.

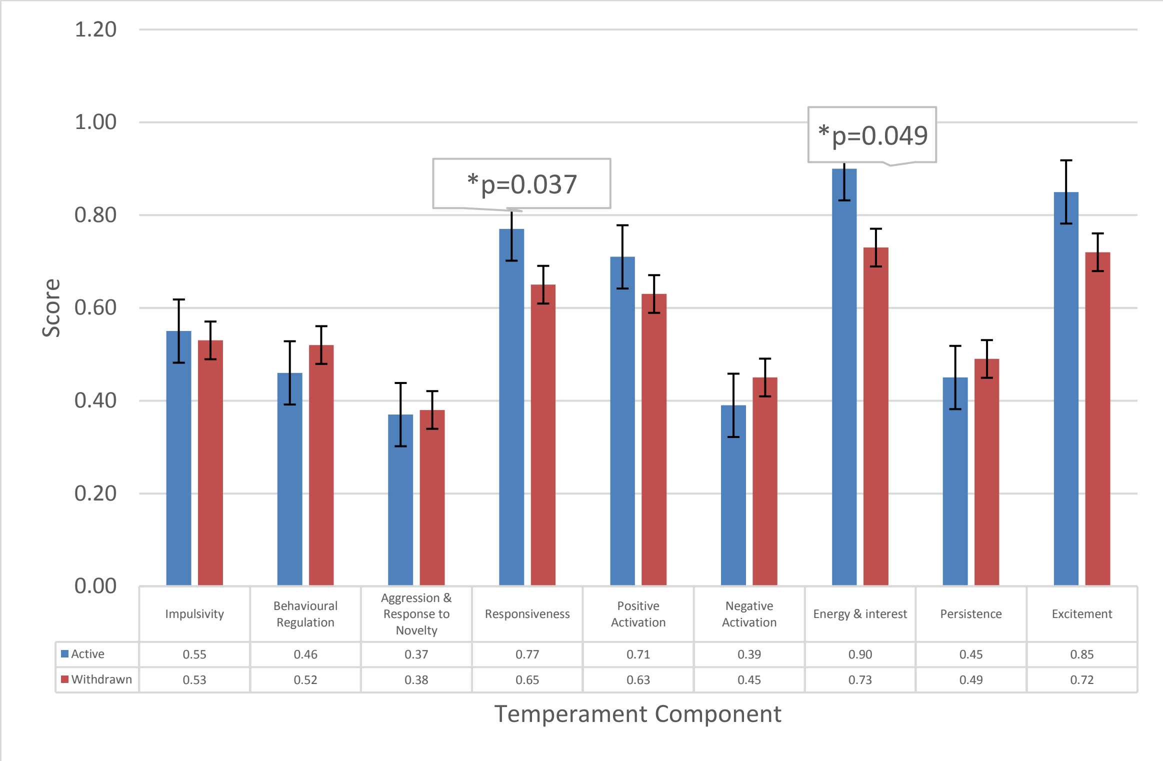


Figure 5.2 A comparison of the temperament components of impulsivity and core affect between active and withdrawn UK police dogs in the matched pairs dataset (n=11 in each population, subjects matched on the basis of age within 1 year, sex, breed, neuter status and whether the dog was general purpose or special purpose).

5.4.5 Comparing Handler Performance Scores for Active and Withdrawn dogs

Independent samples tests were used to search for statistical differences between field score in active and withdrawn dogs (Field score in active (N=55) vs withdrawn (N=12) dogs, $t=5.53$, $p<0.001$), withdrawn dogs having a lower mean field score compared to active dogs. A similar trend was observed in training score between active and withdrawn dogs (Training Score in active (N=55) vs withdrawn (N=12) dogs, ($t=3.36$, $p=0.006$), withdrawn dogs having a lower mean training score compared to active dogs.

Paired samples t tests were used to compare training score and field score for active dogs, withdrawn dogs, and the entire population. All groupings came out scoring statistically lower in the field compared to in training (see Figure 5.3). Significant differences were observed between training and field scores in the active dog group, the withdrawn dog group, and when combining active and withdrawn populations, see Figure 5.4 for active and withdrawn scores graphical representation.

	Active N=55	Withdrawn N=12	Combined N=67
Training Score	4.71 +/- 0.54	3.15 +/-1.31	4.52 +/- 0.84
Field Score	4.27 +/- 0.69	2.58 +/- 1.17	4.07 +/- 1.05
Test for SD between training and field score	.t=4.72 (paired samples t test) P<0.001	.t=2.6 (paired samples t test) P<0.025	t = -4.564 (paired samples t test) P<0.001

Figure 5.3. Mean scores for training and field score in the active, withdrawn, and combined active and withdrawn dogs, and paired samples test results for the differences between training and field scores in each group.

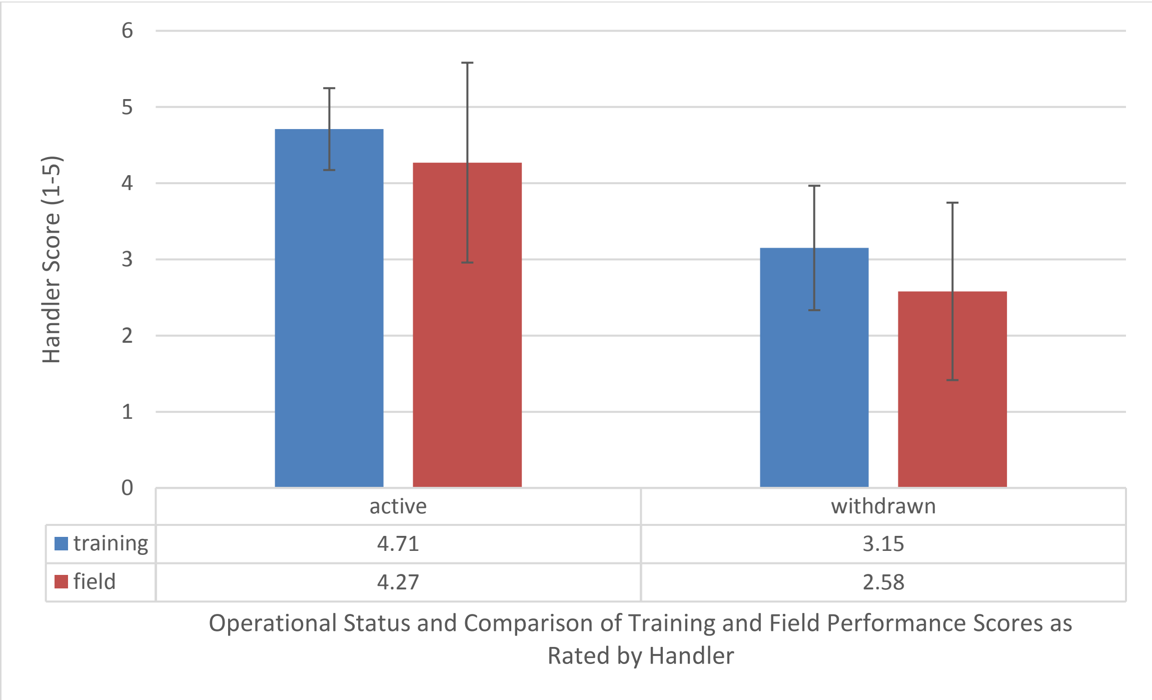


Figure 5.4. Mean scores for training and field score in the active and withdrawn dogs.

5.4.6 Comparisons of Field Scores with PANAS and DIAS Scores

Pearsons Correlations were used to look for relationships between field score as rated by the handler and questionnaire scores on the DIAS and PANAS. See Figure 5.5.

Element of PANAS or DIAS	All dogs Mean, +/- Standard Deviation for questionnaire element	Correlation of Element to Handler Rated Field Score (Pearsons Correlation, p) N=67
DIAS Overall Questionnaire Score (OQS)	0.53 +/-0.07	-0.148, 0.231
DIAS Behavioural Regulation (Factor 1(F1))	0.49 +/-0.13	-0.205, 0.096
DIAS Aggression & Response o Novelty (Factor 2 (F2))	0.53 +/-0.11	-0.162, 0.191
DIAS behavioural Responsiveness (Factor 3 (F3))	0.74 +/-0.11	0.191, 0.122
PANAS Positive Activation	0.72 +/-0.10	0.080, 0.519
PANAS Negative Activation	0.38 +/-0.13	-0.296*, 0.015
PANAS Energy & Interest	0.89 +/-0.12	0.231, 0.060
PANAS Persistence	0.51 +/-0.14	-0.027, 0.826
PANAS Excitement	0.78 +/-0.16	-0.078, 0.529

Figure 5.5. Relationships between handler rated field score for all dogs in the sample and the questionnaire results broken into the elements of the PANAs and the DIAS.

A significant relationship was found between PANAS Negative Activation Score and the Score given for field performance by the handler, as negative activation increased, performance decreased (-0.296*, 0.015). See Figure 5.6.

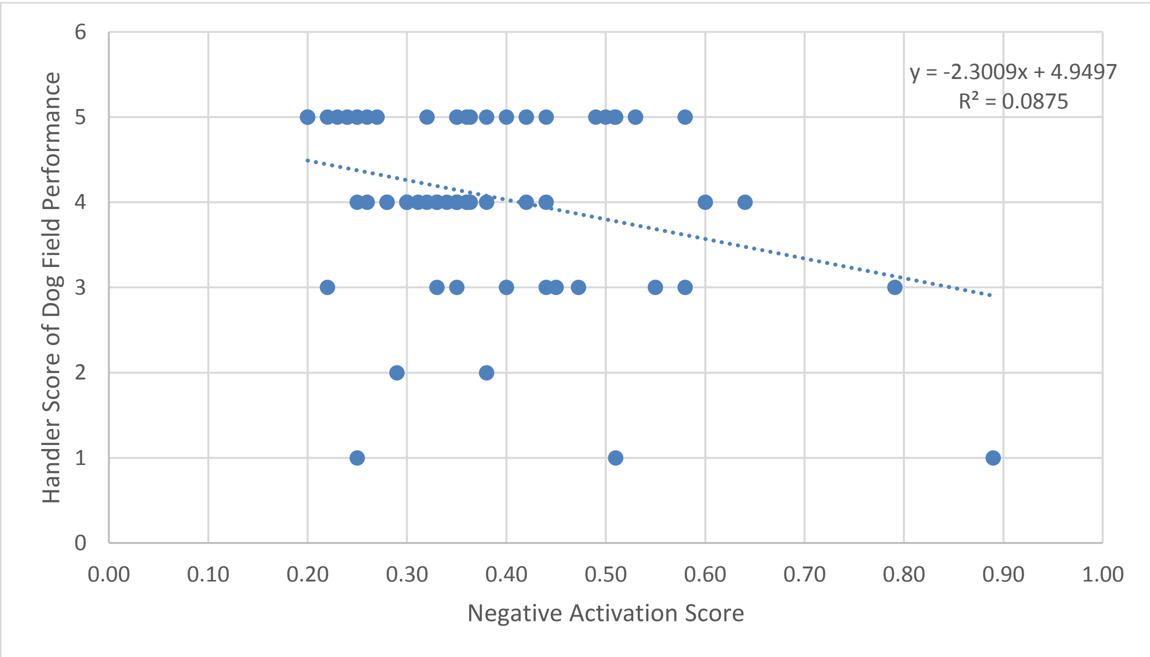


Figure 5.6. Illustration of the direction of the relationship between Negative Activation score from the PANAS and handler allocated field performance score for the working dogs.

5.4.7 Active military and active police dog sample:

No significant differences were found between the active police and active military working dog populations scores (see Table 5.7 below). Effect sizes, particularly for the components of interest were very small, indicating there are likely only minimal differences between the populations as a whole, despite the small sample size in the military population (Figure 5.7 below). The withdrawn dogs could not be compared due to a lack of data on withdrawn military working dogs.

Active Working Dogs comparison				Test for significant difference	Effect size
	component	Military (n=20)	Police (n=63)	Independent t test (t, p)	Hedges G
PANAS	positive activation	0.74	0.74	-0.291, 0.77	0.08
	negative activation	0.39	0.38	-0.380, 0.71	0.10
	energy & interest	0.93	0.92	-0.369, 0.71	0.09
	persistence	0.52	0.52	0.201, 0.84	0.05
	excitement	0.80	0.80	0.125, 0.9	0.03
DIAS	Overall Questionnaire Score (OQS)	0.54	0.55	0.742, 0.46	0.19
	Factor 1: Behavioural Regulation	0.49	0.51	0.641, 0.52	0.16
	Factor 2: Aggression & Response to Novelty	0.34	0.37	0.863, 0.39	0.22
	Factor 3: Behavioural Responsiveness	0.75	0.76	0.148, 0.88	0.05

Figure 5.7 Comparison of Active UK police dog dataset and Active UK military dog dataset in terms of the components of PANAS and DIAS, including effect sizes.

5.4.8 Dutch Dog Data:

The data from the Dutch police dog population only included dogs in active service, so comparisons were made between Dutch active police dogs and UK active police dogs to determine if the populations were similar (Figure 5.8).

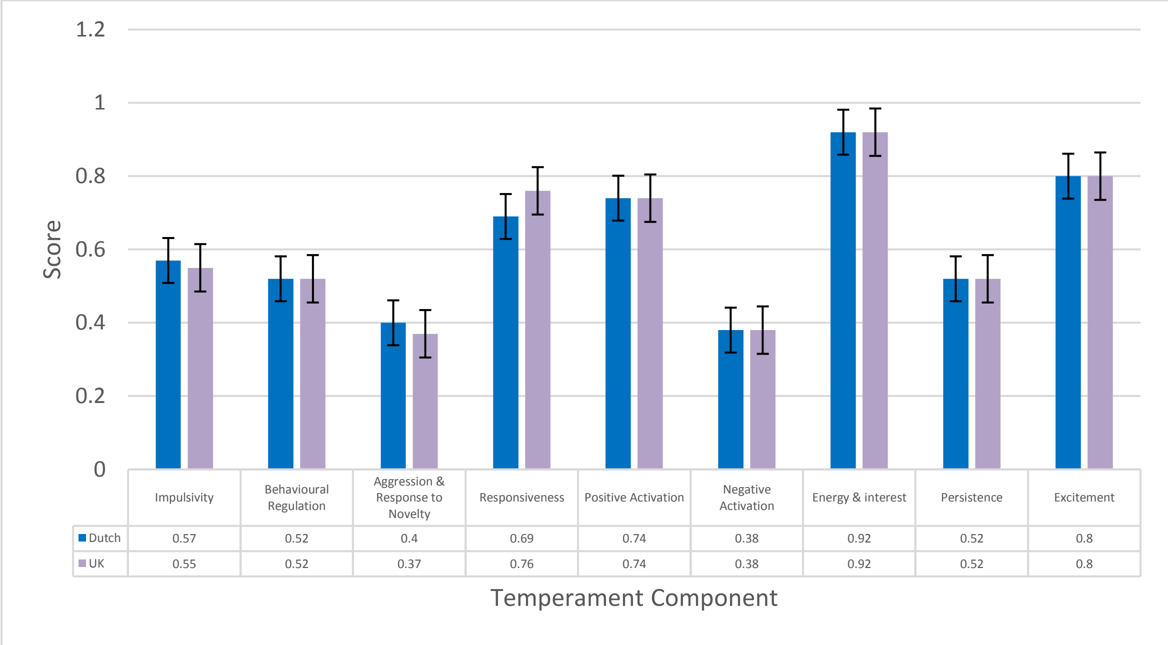


Figure 5.8. A comparison of the temperament components of impulsivity and core affect between active UK police dogs and active Dutch police dogs.

Populations were generally similar (see Figure 5.9 below) with small effect sizes where there was no difference and only a significant difference observed between scores for “Responsiveness” (Factor 3 from the DIAS) between active Dutch dogs (0.69 +/- 0.09) and active UK dogs (0.77 +/- 0.09) (t=0.506, p=0.005).

		Dutch (n=28) Mean score	UK (n=63) Mean score	Independent t test result (t, p)	Hedges G
DIAS Impulsivity	Overall Questionnaire Score (OQS)	0.57	0.55	0.506, 0.53	0.06
	Factor 1: Behavioural Regulation	0.52	0.52	0.212, 0.99	0.00
	Factor 2: Aggression & Response to Novelty	0.40	0.37	0.126, 0.24	0.26
	Factor 3: Behavioural Responsiveness	0.69	0.77	**0.506, 0.005**	0.65
PANAS	Positive Activation	0.73	0.74	0.212, 0.66	0.09
	Negative Activation	0.35	0.38	0.156, 0.29	0.26
	Energy & Interest	0.92	0.92	0.367, 0.81	0.00
	Persistence	0.52	0.52	0.786, 0.87	0.00
	Excitement	0.74	0.8	0.492, 0.11	0.24

Figure 5.9. Comparison of Active UK police dog dataset and Active Dutch dog dataset in terms of the components of PANAS and DIAS, including effect sizes.

5.4.9 Comparisons with the Pet Dog Population

Additional comparisons were made using data collected in previous research (Wright *et al*, 2012) and the UK active working police dog population and the UK withdrawn police dog population. Significant differences were found between the pet dog population and the UK active police dog population for "Behavioural Regulation" (t=2.657, p=0.009), "Responsiveness" (t=2.697, p=0.0082), "Negative Activation" (t=3.759, $p<0.001$) and "Energy & Interest" (t=2.975, p=0.0037). Pet dogs were not significantly different to withdrawn dogs in any trait. See Figure 5.10 below.

	Mean Scores			Pet and Active: independent t-test			Pet and Withdrawn: independent t-test		
Component	Pet (n=560)	Active (n=63)	Withdrawn (n=16)	2 tailed p	t test	SED	2 tailed p	t test	SED
DIAS: Overall Questionnaire Score (OQS)	0.52	0.55	0.56	0.115	1.58	0.01	0.27	1.10	0.03
DIAS Factor 1: Behavioural Regulation	0.47	0.52	0.56	0.009	2.66	0.03	0.05	1.96	0.04
DIAS Factor 2:Aggression & Response to Novelty	0.37	0.37	0.40	1	0	0.03	0.5	0.67	0.04
DIAS Factor 3: Responsiveness	0.699	0.76	0.67	0.0082	2.7	0.02	0.42	0.80	0.03
PANAS: Positive Activation	0.72	0.74	0.68	0.397	0.85	0.02	0.37	0.88	0.04
PANAS: Negative Activation	0.48	0.38	0.45	0.0003	3.76	0.03	0.53	0.62	0.04
PANAS: Energy & interest	0.85	0.92	0.78	0.003	2.98	0.02	0.17	0.18	0.05
PANAS: Persistence	0.55	0.52	0.54	0.342	0.96	0.03	0.84	0.19	0.05
PANAS: Excitement	0.79	0.80	0.77	0.753	0.31	0.03	0.71	0.37	0.05

Figure 5.10. Comparison of PANAS and DIAS component scores for pet dogs, active UK police dogs and withdrawn UK police dogs using independent samples t tests.

5.5 Discussion:

The aim of this work was to identify components of temperament based around core affect and impulsivity that are associated with withdrawal from service of working dogs in the United Kingdom.

Dogs that had been withdrawn from service had significantly lower scores for two other traits compared to dogs in active service (based on the group data: “Responsiveness” p=0.008, “Energy & Interest” p=0.02), and were more similar to the pet dog population in this regard (“Responsiveness” p=0.425, “Energy & Interest “p=0.172). This suggests that working dogs are different in terms of these traits and these traits are potentially linked to staying in active service. It needs to be borne in mind that without a longitudinal study it cannot be determined if, dogs are withdrawn because they possess this profile as part of their temperament at selection, or if dogs that are failing to perform at work develop this profile as they lose motivation to work.

Expectedly, withdrawn dogs were scored lower in terms of performance than active dogs both in training and in the field by their handlers, as they had been withdrawn from service this was to be expected. However, it is important to note that handlers were asked to score their dogs up to one year after withdrawal had taken place, so the fact their dog had been withdrawn may have resulted in the lower score rather than recall of the dogs actual performance. With any information collected through retrospective recall there could be issues with accurate memory of the information which could be an issue in this study, although limiting the withdrawal period to within a year was used as a counter measure for this.

Both active and withdrawn dogs were scored lower for performance in the field compared to in training. This is potentially because training is still a controlled environment, whereas performance in the field is more subjective sometimes as the handler won't always know if the dog has found every item they were trained to detect, or if they followed the track accurately. This unknown could result in handlers scoring their dogs performance more conservatively in the field.

As negative activation increased, field performance score decreased (-0.296^{*}, 0.015). This ties into the literature on fear as a factor relating to failure in working dogs (Murphy, 1995. Ruefenacht *et al*, 2002. Batt et al, 2008. Gosling et al, 2009. Caron-Lormier *et al*, 2016. Rooney *et al*, 2016. Ilska *et al*, 2017. See Chapter 1 for full discussion) as negative activation as scored on the PANAS taps into fear related behaviour in terms of sensitivity to negative qualities in the environment (Shepherd & Mills, 2006). Potentially scoring the dogs and looking at Negative Activation could indicate how well the dog is performing in the field, or how well the dog would be predicted to perform in the field, more research would be needed to track dog's scores to establish if the relationship is cause or effect. This was the opposite effect to what was observed for the search dogs in Chapter 2 in terms of performance and negative activation, which higher levels resulted in better overall performance. However this was scored objectively in terms of the number of finds, rather than asking the handler to score their own dog's performance, so differences in scoring methods and handler interpretation of dog performance could impact on the results and handler scores may not accurately reflect performance. The handler score for the dog is important though, as it gives an indication of how the handler perceives the dog, dogs perceived to be performing well are likely to be rewarded more and have a more positive relationship with their handler, which could contribute to success, and the opposite could apply to those perceived to be performing less well.

UK military working dogs in active service were found to have a statistically similar profile to UK police dogs in active service. Particularly noting scores for “Energy & Interest” (p=0.71) and for “Responsiveness” (p=0.88), which are the factors that differed between active and withdrawn dogs. However conclusions need to be drawn with caution due to the limited sample size. Ideally additional data would be needed to state with confidence that the same factors related to withdrawal in police dogs also apply to military dogs, although small effect sizes indicate that the samples are likely to be similar.

The active Dutch police dog dataset was remarkably similar to the UK active police dog data, except in terms of “Responsiveness” where the Dutch population scored significantly lower (p=0.005). In this respect the Dutch active dogs (0.69) more closely resembled the UK pet dog population (0.69), with the UK withdrawn dogs still scoring lower for this trait (0.65). There are a number of possible explanations for this finding; it could be that there is not a positive selection for this trait component in the country, perhaps due to different working requirements; potentially what makes a successful working dog varies depending on the country; perhaps population norms differ between the countries and thus influence the available selection pool; it could be a result of random error; training factors might have influenced this; or finally it could be that the questionnaire is answered and interpreted differently in a different country. Wan *et al*, 2009 identified cultural differences in the rating of German Shepherd dogs between the USA and Hungary, which while the results could have been due to dogs being genetically different because of their origin, it could be the result of cultural differences in the interpretation of dog behaviour. Without targeted information regarding each of these explanations the finding is difficult to explain further. These hypotheses could however be explored in future work. While the dogs were statistically similar to UK working dogs in terms of their levels of “Energy & Interest”, perhaps “Responsiveness” has more to do with how the handlers train and engage the dogs and could indicate that such dogs could be suitable for work if the right motivators can be found to make them more responsive. This might reflect cultural differences in the training of dogs and thus the importance of this trait. It could also indicate that “Energy & Interest” as a trait plays more of a role in whether a dog is withdrawn, or is more effected as a trait by dogs as they start to fail in work.

Similar traits to those identified as being related to dogs in active service in the present study have previously been linked to successful training outcomes in drug detection dogs, where desire for work and distractibility were linked to success in terms of completing training (Maejima *et al*, 2007). Engagement, in terms of physical and social interactions with people and objects, has also been linked to certification in military working dogs (Foyer *et al*, 2015). These traits could be linked to “Responsiveness” and “Energy & Interest” that were found to be related to field success in the present study. While previous research has examined the relationships between temperament and training success or certification in working dogs, the present study focused on temperament traits related to continued success in the field. If these factors are comparable then it is possible that some of the tests predicting certification may also predict ongoing field success, but this hypothesis would need to be tested by longitudinal studies following up certified dogs in the field and studies correlating the PANAS and DIAS scores with those of the certification tests.

The other components of the temperament traits measured using the PANAS and DIAS were not associated with being withdrawn. These included "Overall Questionnaire Score for Impulsivity," "Behavioural Regulation," and "Aggression and Response to Novelty" from the DIAS, and "Negative Activation", "Positive Activation," "Excitement" and "Persistence" from the PANAS. These traits may simply not be related to working dog success even if the working population does differ from the pet population in some of them, as in previous studies not all temperament traits measured have been linked to performance outcomes (Seibert & Kraimer, 2001). Or it might be that they do not readily change or cause a problem in the field.

UK withdrawn police dogs do not appear to be different to the pet dog population studied in terms of any of the factors of core affect or impulsivity. However, the pet dog population (used for norm referencing from Wright *et al*, 2012) significantly differed to the UK active police dog sample with significantly lower scores for the elements 'Behavioural Regulation' (active=0.52, pet=0.47, $p=0.009$) and higher scores than the UK active police dog sample in terms of 'Negative Activation' (active=0.38, pet=0.48, $p<0.001$) of core affect from the PANAS. This suggests that there is positive selection for certain qualities in working dogs and makes it unlikely that these traits play no role in terms of withdrawal. Perhaps their impact on performance is more complicated and varied in different working roles (from various types of detection work to protection work) making it difficult to pull apart differences in the working dog population as a whole. This area needs further clarification specifically through the comparison of subsets of the working dog population.

The low negative activation observed in UK active police dogs when compared with the pet population is potentially very important as this would seem to reflect them being less affected by a stressful and potentially intimidating working environment. As assessing fearfulness is a focus of most selection tests this finding bears out its likely usefulness as it is likely to tap into negative activation. However, while fear is recognised as an important trait in terms of success (see Chapter 1) it is important to remember that this is not the only factor associated with a dog's success, although this research further supports the notion that sensitivity to potentially aversive or fear eliciting stimuli is of relevance and can also be measured using the PANAS.

In humans it has been found that a number of processes are impacted on by emotional states. The valence of the emotion has been found to influence attention, memory and judgement. Negative states such as anxiety change attention bias to potentially threatening stimuli (e.g. Mathews and Macleod, 1994; Mineka *et al*, 1998; Mogg and Bradley, 1998), while positive states such as happiness improves memory recall (e.g. Bower, 1981; Burke & Mathews, 1992; Denny & Hunt, 1992; Mineka *et al*, 1998). It has also been shown that people in positive emotional states are more optimistic about the future, while those in negative emotional states are more pessimistic (e.g. Eysenck *et al*, 1991; Wright & Bower, 1992; MacLeod & Byrne, 1996; Nygren *et al*, 1996). While it is acknowledged that different levels of emotional valence within either the positive or negative spectrum can result in different influences on cognitive function (Lerner & Keltner, 2000), there is the acknowledgement that traits may impact on the expression of and predisposition towards positive or negative states (Mineka *et al*, 1998; Mogg & Bradley, 2005). This is where core affect as measured by the PANAS could link in as a method of measuring predispositions towards positive and negative attentional bias. While this is in humans, parallels have been drawn to animals (e.g. Clark & Squire, 1998; Call & Carpenter, 2001; Desiré *et al*. 2002; Garner *et al*, 2003; Smith *et al*, 2003; Paul *et al*, 2005; Herry *et al*, 2007; Shafir *et al*, 2008) and research into cognitive bias in dogs has revealed differences in perception of ambiguous stimuli following events to elicit different emotional states (Burman *et al*, 2011). This is where the PANAS could be a useful tool to examine such predispositions, and was the focus in this thesis to look at underlying mechanisms relating to the predisposition of individuals in terms of how sensitive they are to positive and negative qualities within the environment. It may be the case that individuals scoring highly for positive activation on the PANAS respond more optimistically in cognitive bias testing despite ambiguous stimuli being presented, and the opposite may be true for those scoring highly for negative activation. This could help us to understand how predispositions to positive and negative valence impact on emotional states and behavioral expressions in the here and now, and the extent to which predispositions effect choices and stimuli perception.

Longitudinal research to track from selection through to outcome of entire working life in a variety of working dog disciplines could provide more specific data on the casual link between temperament traits associated with withdrawal. Sample size in general for this study was relatively small compared to a number of the studies identified in Chapter 3, potentially due to the volume of information being requested from handlers being too time consuming for completion. The project was also limited by the sponsors and required clearance to approach dog units and departments which limited the sample size. Due to the small sample size it was not possible to look at the dogs in terms of the working role they were performing and analysing differences in this in terms of withdrawal. While existing temperament tests often have a focus on fear behaviour, depending on the role the dog is going to perform will depend on the selection testing focus (as discussed in Chapter 3). So while in police and military dogs there is a focus on being quick to act and showing an element of stranger directed aggression (as discussed in Chapter 3), for guide dogs and hearing dogs the focus is more on being calm and biddable (as discussed in Chapter 3). While withdrawal may result from different factors based on the dogs working role, general overarching factors such as fear and a lack of willingness to work would apply to withdrawal regardless of role, the focus of this project was to look for such overarching factors.

While validated questionnaires can be a useful tool for assessing temperament traits in working dogs in this study and have been used to predict guide dog qualification with 80% accuracy (Arata *et al* 2010), they require the dog's caregiver to be able to answer the questions reliably and truthfully. This is not always possible, particularly when dogs are sourced from private companies or rescue centres. This means that to be of real value to the working dog industry, such traits need to show convergent validity with simple and effective behaviour tests to enable dogs to be assessed regardless of their source or situation. If the relationship is not causal but the traits change over time as a result of experiences in training and work, the questionnaire can act as a useful tracking tool to monitor dogs. Having identified components that might be associated with withdrawal the next question is can we devise a simpler test that does not depend on knowledge to get a result. The PANAS and DIAS are believed to be reliable and valid tools to assess temperament traits in pet dogs (Taylor & mills, 2006. Wright *et al*, 2012) and this study shows they have potential value for use in assisting with assessing working dogs as well. As they have been shown to be reliable measures (Wright *et al*, 2012), this means that it should be possible to develop behaviour tests around the relevant traits and have the potential to be used to predict individuals at risk of withdrawal from service.

5.6 Conclusions:

Temperament traits ("Responsiveness" in terms of impulsivity from the DIAS, and "Energy & Interest" in terms of positive activation from the PANAS) were found to differ between dogs in active service and those withdrawn from service in a population of UK police dogs. Further work is needed to determine cause or effect in relation to these traits. If the relationship is causal, factors relating to withdrawal of dogs from service can potentially be used to predict which dogs are unsuitable for work, or at risk of being withdrawn in the field before selection. It could also be used to identify individuals currently working that are at risk of withdrawal with the aim to implement remedial interventions. With an understanding of factors that put dogs at risk of withdrawal, there is a need to develop a way to predict these factors or track their development to further facilitate selection testing of working dogs. Questionnaires cannot always be used or relied on in working dog selection as they require completion by someone with a level of knowledge to reliably answer the questions without a vested interest. There is therefore a need to develop behaviour tests to measure these traits so that the traits can be measured in a practical way, for example before selection by people who do not know the dog and thus cannot complete a psychometric assessment.

Chapter 6: Using a Pet Dog Population to Develop Behaviour Tests to Measure Reported Traits of ‘Responsiveness’ and ‘Energy & Interest’

Synopsis

Based on the findings of the previous chapter, behaviour tests were developed to measure components of the trait “Responsiveness” (in terms of positive activation based on the PANAS scale), and “Energy & Interest” (in terms of impulsivity based on the DIAS scale). The study recruited 20 pet dogs: 10 of which closely matched the score profiles of working police dogs in active service from the previous study in Chapter 5, the other ten had score profiles matching withdrawn working dog profiles (in terms of the “Responsiveness” component of impulsivity and the “Energy & Interest” component of positive activation). These were then defined as the high scoring group (those matching the in active service profile), and the low scoring group (those matching the withdrawn working dog profile. Of the six tests developed, results from one of the tests identified significant differences associated with group membership. There was a significant difference between the higher scoring and lower scoring groups in terms of the number of trials taken to learn the discrimination task at test (p=0.021) with dogs in the higher scoring group requiring fewer trials to meet learning criteria, but this difference was not observed at retest (0.46). The lower scoring group learned the discrimination task in significantly fewer trials in the retest (average 31.8 trials) compared to their performance in the test situation (average 88.6 trials, p=0.009). Differences between test and retest results between the groups might indicate either differences in rates of learning, memory or recall between them or unreliability of the test. The results suggest that behavioural differences can be observed between the groups of dogs, compared to those with the lower scoring profile, those with the higher scoring profile responded to the test in a way that indicates faster task acquisition, which may be an important attribute for long term success, as they learn about and potentially adapt to their environment more quickly. A prospective study is needed to establish population norms for performance within these tests and whether these tests could be a useful addition to working dog selection procedures.

6.1 Introduction

Working dogs need to be able to cope in a variety of settings and situations (Foyer *et al*, 2015), and, in the absence of objective methods of assessment, there is a danger that selection may be biased by the opinion of what is preferred by test leaders, and not necessarily what would objectively be a successful working dog in the field (Foyer *et al*, 2015). What constitutes a ‘good working dog’, likely varies across disciplines and regions, but generally means the dog is able to work effectively in the field, despite challenging working conditions (Foyer *et al*, 2015). Currently selection is largely subjective (Foyer *et al*, 2015) which may also explain why withdrawal from field service in working dogs is a problem. It has been suggested in one study that eighty two percent of dogs over the age of 5 that are withdrawn from active service are withdrawn due to behavioural problems (Evans *et al*, 2007). This is despite the use of selection tests, suggesting that there is still room for improvement when it comes to identifying factors associated with success or withdrawal from service.

The previous study (Chapter 5) examined a limited range of temperament traits linked to withdrawal from service in working dogs using two psychometric questionnaires. The temperament profiles generated from this study identified two temperament traits of relevance in terms of dogs being withdrawn from service in the population of police working dogs, one related to a facet of impulsivity described within the DIAS (Factor 3, Responsiveness) and one within the positive component of PANAS (Positive Activation subcomponent "Energy & Interest"). Dogs that score highly for "Responsiveness" are thought to be easier to train, remain interested in new stimuli for longer, and react more quickly (Wright et al, 2011). Having high levels of "Responsiveness" is generally thought of as being desirable when training a dog (Wright et al, 2011). By contrast, "Energy & Interest" relates to the dog's general level of vigour, and engagement they show in new tasks, environments or stimuli (Taylor & Mills, 2006).

While the PANAS and DIAS have been shown to be reliable and valid tools for measuring temperament traits in pet dogs and might potentially be of use to score working dogs to identify individuals at risk of withdrawal (assuming the correlational relationship observed in Chapter 5 is causal and not consequential), questionnaires are limited in their usability as they require a level of knowledge about the dog which is not always possible given the way that working dogs are sourced. In the working dog sector having an additional measure such as a behaviour test can provide a tool to be used when there is not a reliable or consistent caregiver able to answer questions accurately about the dog. Behavioural test correlates might also give more specific insight into the nature of any possible causal relationship between behavioural tendency and performance in the field.

It was found previously that the spatial discounting paradigm developed to assess impulsivity in general is robust enough to be simplified so that it is feasible for field use (See Chapter 4). The test itself only measures the broader trait impulsivity, though research from Chapter 5 suggests that the way in which impulsivity relates to field performance is more complicated than this, meaning tests need to be developed to measure the components within impulsivity. In the current chapter the aim was to develop a battery of feasible behaviour tests for potential use by working dog selectors which further the ability to predict both aspects of impulsivity and aspects of positive activation levels which are correlated with withdrawal from service later in life.

From Chapter 5, "Responsiveness" from the DIAS, and "Energy & Interest" from the PANAS were identified as potentially important. These components are made up of a number of questionnaire elements. "Responsiveness" questionnaire elements making up this component include 'My dog appears to be 'sorry' after it has done something wrong', 'My dog is easy to train', 'My dog takes a long time to lose interest in new things', 'My dog is very interested in new things and new places', 'My dog reacts very quickly'. The second component of interest, Positive Activation component "Energy & Interest" in terms of core affect (using the PANAS-Positive and Negative Activation Scale) significantly differed between working and active dog population, with active dogs scoring higher. PANAS, "Energy & Interest" questionnaire components include, 'Your dog is full of energy', 'Your dog shows little interest in its surroundings' (reverse scored), 'Your dog is lazy' (reverse scored), 'Your dog requires a great deal of encouragement to take part in energetic activities' (reverse scored). Where reverse score is indicated the opposite of the statement applies.

The tests developed within this part of the project were devised by a panel of animal behaviour experts at the University of Lincoln and based on existing tests in the literature that examined similar concepts to those of interest within this project, but were developed to be used with dogs in a quick assessment. The basis of the tests, rationale for their development and proposed relationship to the elements from the PANAS and DIAS are outlined below.

Novel or unfamiliar environments that are complex offer opportunities for widespread investigation, such tests usually allow a dog to investigate a novel room filled with objects to assess exploratory behaviour and confidence in working dogs (Sherman *et al*. 2015. Svartberg, 2002. Svartberg, 2005. Svartberg *et al*. 2005. Tomkins *et al*. 2011). In terms of the component "Energy & Interest", dogs that are more interested in their environment can be expected to spend more time investigating new environments to gather information they desire compared to individuals that are less interested. This is based on the elements assessed within "Responsiveness", 'my dog takes a long time to lose interest in new things' and 'my dog is very interested in new things and new places'. In addition to this time spent investigating a novel environment might correlate with the "Energy & Interest" component of positive activation with the elements 'your dog shows little interest in its surroundings' (reverse scored so the opposite applies, meaning the dog shows interest in the surroundings so is more likely to investigate), 'your dog is full of energy' and 'your dog is lazy' (reverse scored), and 'your dog requires a great deal of encouragement to take part in energetic activities' (reverse scored so the dog requires little encouragement, potentially making them more willing to investigate without encouragement). A novel environment test is readily used in working dog selection tests (Sherman *et al*. 2015. Svartberg, 2002. Svartberg, 2005. Svartberg *et al*. 2005. Tomkins *et al*. 2011. Observations at Greater Manchester Police Dog Unit, 2014.), identifying relationships between test variables and the components linked to withdrawal in

Chapter 5 could provide additional information for selection procedures. As discussed such tests are often used as part of assessment of confidence, but may actually provide information on a number of other elements as well as confidence (such as those being tested within this thesis), that would be of value to be aware of in working dogs.

Novel object tests assess the degree to which an individual engages with an unfamiliar object, they are widely used to assess fearfulness (Ley *et al*, 2007) which is acknowledged as an important trait in terms of working dog failure (Murphy, 1995. Ruefenacht *et al*, 2002. Batt *et al*, 2008. Caron-Lormier *et al*, 2016. Ilska *et al*, 2017). Adaptations of this, such as a two part test assesses engagement with a specific change in the environment and has been used for decades in the assessment of behaviour in rats (Bindra & Spinner, 1958), as well as in dogs to assess response to novelty (Ley *et al*, 2007). Such tests can be used to assess fear and environmental engagement in animals by recording and analysing their behaviour change when a change occurs within the environment (Bindra & Spinner, 1958). Dogs that are more interested (as indicated by the elements 'my dog takes a long time to lose interest in new things' and 'my dog is very interested in new things and new places') in terms of the "Responsiveness" component of impulsivity, and more sensitive to their environment in terms of the "Energy & Interest" component of positive activation, 'your dog shows little interest in its surroundings' (this item is reverse scored so the opposite applies) may be expected to be more likely to notice when things change and seek information about the change through investigation. Being aware of changing environments is an important attribute for working dogs, while awareness is necessary, distraction because of change can cause problems (Maejima *et al*, 2007).

Attention is defined as taking notice of something and implies a level of awareness about one's surroundings. Dogs that have higher levels of "Energy & Interest" (in terms of the "Energy & Interest" component of positive activation, in particular the elements 'your dog shows little interest in its surroundings' (reverse scored), 'your dog is full of energy', your dog is lazy' (reverse scored), 'your dog requires a great deal of encouragement to take part in energetic activities' (reverse scored)), and dogs which are more responsive (in terms of the "Responsiveness" component of impulsivity, and in particular the elements measured within this component 'my dog takes a long time to lose interest in new things', my dog is very interested in new things and new places', and 'my dog reacts very quickly'), may be more likely to attend to change as they are potentially more aware of what is going on around them. Thus they might be expected to perform better in a visual tracking task built on attention to change within a routine.

A discrimination task involves an individual learning to identify the difference between two stimuli, usually by learning that interacting with one stimulus results in a reward while the other does not.

Rooney & Cowan, 2011 tested discrimination in dogs through training them to target one marked wooden spoon over another, as part of a project testing the impact of previous training history on performance. Dogs that score higher in terms of "Responsiveness" from the DIAS (particularly in relation to the element 'my dog is easy to train' within the "Energy & Interest" component), are likely to acquire new tasks such as a discrimination test, faster than lower scoring individuals. In terms of the "Energy & Interest" component the elements within this components likely to be tapped into by this test are 'your dog is full of energy' and 'your dog is lazy' (reverse scored), as this may make them more likely to engage with a task, and the element 'your dog requires a great deal of encouragement to take part in energetic activities' (reverse scored), as this impacts on a dogs level of engagement with a task. Dogs scoring higher for this element are likely to show greater engagement with the task and be easier to train, again suggesting that they would acquire the task faster.

Working dog tests of startle reactions and response to acoustic stimuli have been used to assess fearful and reactive behaviour in dogs by examining their reactions as well as their latency to recover after the unexpected noise (Batt *et al*. 2008. Tomkins *et al.* 2011). Such tests allow the dog to investigate a room, and then produce a startle noise, anything from shouting, to a metallic noise, to a gunshot. (Asher *et al*. 2013. Sherman *et al*. 2015. Svartberg, 2002. Svartberg, 2005. Svartberg *et al*. 2005, Svobodova *et al*. 2008). The test developed within this project to assess social sensitivity was based around these already in use in the working dog sector, but utilised a context more frequently seen in a pet dog home (being reprimanded for 'bad' behaviour), to meet ethical guidelines. Social sensitivity includes the ability to perceive and respond appropriately to the emotions of others, and in the test developed in this project this was tested through verbal admonishment while investigating an area. Dogs that score higher for "Responsiveness" are likely to respond more to the socially sensitive situation within this test if the differences relates to the element 'my dog appears to be 'sorry' after he has done something wrong', which suggests an appropriate response to the social situation.

Following on from this, social cues are verbal or non-verbal indicators, tests in animal studies often use hand signals or pointing gestures as a social cue. Ability to follow pointing gestures has been assessed in dogs and has highlighted that provided the dogs have had sufficient social experience with humans they are able to follow human pointing gestures (Lazarowski, et al 2015). The ability to then reverse such learning has also been tested in dogs (Milgram *et al*, 2004) and is thought to test flexibility and ability to try alternative options when what previously worked no longer pays off (Milgram *et al*, 2004). The reversal learning task developed in this project was based on Milgram *et al*, 2004 which used reversal learning tasks to assess age related learning in beagle dogs where they were taught to avoid an object previously associated with reward and respond to an object that had previously no history of reward (Milgram *et al*, 2004). Dogs that are more responsive in terms of the elements within "Energy & Interest" - 'your dog is full of energy', 'your dog is lazy' (reverse scored), and 'your dog requires a great deal of encouragement to take part in energetic activities' (reverse scored), may be more likely to engage with and persist with the task of learning and then learning to ignore social cues. Dogs that score higher for "Responsiveness", are likely to be more responsive to social cues if the difference is based on the element 'my dog appears to be 'sorry' after it has done something wrong', and learn the task faster if it is based on the elements 'my dog is easy to train', and 'my dog reacts very quickly'.

6.1.1 Aim:

A test battery was designed to try to determine whether it is possible to discriminate between dogs with the traits found to be correlated with withdrawal from service ("Responsiveness" and "Energy & Interest") from the questionnaires distributed to the police dog population in Chapter 5. This would negate the need for the detailed knowledge of the dogs which a survey requires.

6.2 Methods:

The present study was based on data collected in Chapter 5 where in police dogs two components of temperament were found to be significantly different between active working dogs and those withdrawn from work. Behaviour tests were designed to map onto these elements as descried in Chapter 6.1. See Figure 6.1 below.

Questionnaire Component	Items mapped on to proposed behaviour tests						
	Items	Novel Environment	Novel Object	Attention	Discrimination	Social Sensitivity	Social Cues
PANAS: Energy & Interest	Your dog shows little interest in its surroundings	✓	✓	✓			
	Your dog is full of energy	✓		✓	✓		✓
	Your dog is lazy	✓		✓	✓		✓
	Your dog requires a great deal of encouragemen t to take part in energetic activities	✓	✓	✓	✓		✓
DIAS: Factor 3 Responsivenes s	My dog appears to be 'sorry' after it has done something wrong					✓	✓
	My dog is easy to train				✓		✓
	My dog takes a long time to lose interest in new things	✓	✓	✓			
	My dog is very interested in new things and new places	✓	✓	✓			
	My dog reacts very quickly			✓			✓

Figure 6.1 Proposed behaviour tests and their proposed relationship to temperament components, based on specific items from the PANAS and DIAS questionnaires that contribute to each component. The tick indicates which elements of the questionnaires are thought to be being examined in each of the designed tests.

6.2.1 Subjects:

Pet dog owners were recruited to complete the PANAS and DIAS questionnaires online through the University of Lincoln Pets Can Do database during the time period between April 2015 and July 2015 (http://www.lincolnpetscando.co.uk/). From these responses 10 dogs were selected which scored in a similar range (closest matches were selected from those that applied to take part to ensure the smallest deviation possible within the time constraints) to active working dogs based on the analysis of the police dog dataset (DIAS Factor 3 (0.77+/-0.05), PANAS Energy & Interest (0.90 +/-0.05)), and 10 dogs were recruited which scored in a similar range to dogs to those that were withdrawn due to behavioural problems (DIAS Factor 3 (0.65+/-0.05), PANAS Energy & Interest (0.73 +/-0.05). See Figure 6.2. The 20 dogs selected for the study were taken from 47 responses to the initial recruitment.

Group	**Breed**	**Age (years)**	**Gender**	**Neutered**
High	Cocker Spaniel	3	F	N
High	Cockapoo	4	M	Y
High	Border Collie	3	M	N
High	Mixed	4	M	Y
High	Wheaton Terrier	1	M	N
High	Labrador	4	M	Y
High	Border Terrier	8	M	Y
High	Labrador	10	F	Y
High	Mixed	6	M	Y
High	Labrador	2	F	N
High	Saluki	1	M	N
Low	Border Collie	8	M	Y
Low	Labrador	10	F	Y
Low	Cocker Spaniel	4	M	Y
Low	Border Collie	1	M	Y
Low	English Springer Spaniel	8	M	Y
Low	Hungarian Viszla	6	F	Y
Low	Lurcher	3	M	Y
Low	Working cocker spaniel	9	M	N
Low	Cairne Terrier	3	M	Y
Low	Mixed	3	F	Y

Figure 6.2. Demographics of the dogs recruited for the study, including their breed, age, sex and neuter status, as well as which group they belonged to based on their scores generated from the PANAS and DIAS questionnaires they completed (high or low scoring, where high scoring dogs were recruited as they had similar scores to dogs from the active police dog population in Chapter 5, and low scoring dogs recruited as they had similar scores to the withdrawn police dog population in Chapter 5).

The dogs selected for testing were then compared in terms of the elements of PANAS and DIAS of interest to ensure that the groups were statistically significantly different (see Figure 6.3).

	higher scoring (n=10)	lower scoring (n=10)	t	P
PANAS: Energy and Interest	0.95 +/- 0.047	0.76 +/- 0.039	9.775	<0.001
DIAS Factor 3: Responsiveness	0.79 +/- 0.034	0.63 +/- 0.032	10.57	<0.001

Figure 6.3. Comparisons of the test population in terms of “Energy & Interest” and “Responsiveness” using independent samples t tests. Yellow highlight indicates a statistically significant differences was identified (based on p<0.05).

The distribution of dogs within each of the study groups was also examined (see Figure 6.4) to show the spread of scores within the test groups and how they differed from the rest of the population that completed the questionnaires for the study.

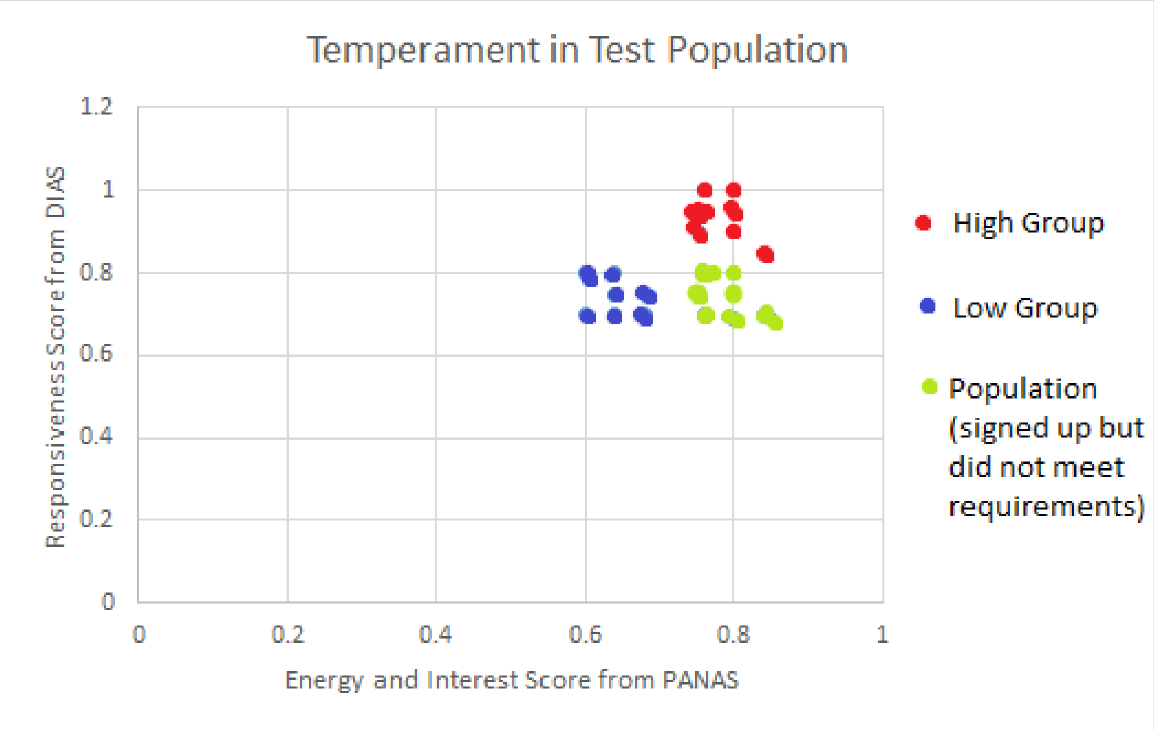


Figure 6.4 Temperament distribution of the study population, selected for closely mapping the results of the previous study (Chapter 5), the higher scoring group matching the operational working dogs and the lower scoring group matching profiles of withdrawn working dogs in terms of the temperament components “Energy & Interest” and “Responsiveness”. The population of dogs that signed up for the study but did not meet the requirements are also included, demonstrating both high and low groups differed from the ‘normal’ population. The groups are clearly separated.

Between each test every dog was given a ten minute break to go outside with their owner, all testing was carried out at the University of Lincoln Minster House, LN6 9DL.

6.2.2 Novel Environment Test:

A selection of novel objects and surfaces were set up in a room measuring 5 metres by 6 metres (see Figure 6.5 for diagram, and Figure 6.6 for photograph of the set up). The room and the objects were cleaned with safe4 disinfectant before testing each dog. The dog was released into the room by the owner through a sliding door which was closed behind the dog with the owner outside of the room and out of sight and hearing. The researcher was positioned behind the video camera (see Figure 6.5) and did not interact with or look at the dog, The dog was given the opportunity to investigate the room until they stopped investigating (investigating was defined by sniffing the objects or the room) for either two blocks of 15 seconds within a 2 minute period or one continuous block of 30 seconds. At this point the tester stopped the camera and opened the doors to allow the dog to leave the room and return to their owner.

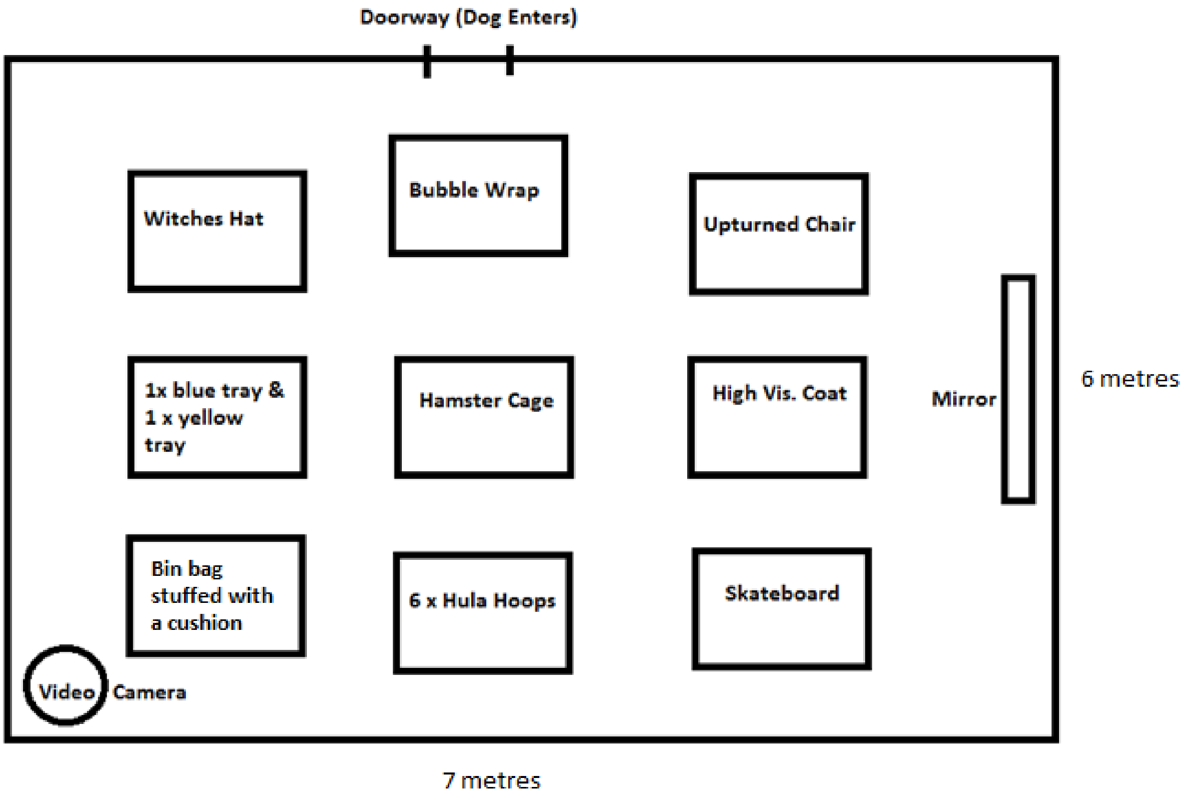


Figure 6.5 Diagram of the room set up including details of placement of novel objects in the novel environment test. The researcher was positioned behind the video camera. Items were all 1metre apart from each other and those bordering a wall were 50cm away from a wall, except the mirror which was centrally placed along the wall as illustrated.

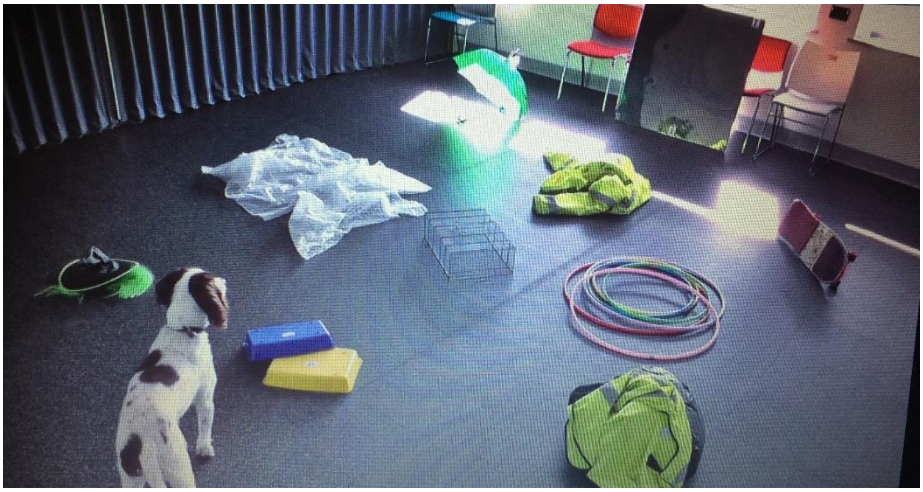


Figure 6.6 Photograph of the Novel Environment test set up. The dog can be seen in the room with the novel objects.

Time spent investigating the objects in the room in the novel environment test could be related with group status (higher scoring or lower scoring), as the novel environment test was designed to measure all of the elements within the "Energy & Interest" component of the PANAS, and 2 of the elements within the "Responsiveness" component of the DIAS. Dogs in the higher scoring group (in terms of the DIAS scores for "Responsiveness" and PANAS scores for "Energy & Interest") were hypothesised to spend significantly longer investigating than dogs in the lower scoring group (in terms of object sniffing time), because higher scores indicate more interest in the environment (PANAS 'your dog shows little interest in its surroundings' (reverse scored), DIAS 'my dog takes a long time to lose interest in new things' and 'my dog is very interested in new things and new places', as well as energy and engagement with activities (PANAS 'your dog is full of energy', 'your dog is lazy' (reverse scored), 'your dog requires a great deal of encouragement to take part in energetic activities' (reverse scored). However, given the generality of components measures, this test might be a relatively insensitive measure of difference between the populations.

6.2.3 Novel Object Test:

The dog was let into the room measuring 6 metres by 7 metres (a secondary room to the novel environment test) through two sliding doors that were parted by the owner and then closed behind the dog once the dog had passed into the test room. Upon entering the room the dog was exposed to 2 identical objects (upturned waste paper bins that had been cleaned with safe 4 disinfectant before each test) that were placed on the floor directly visible from the dogs entry point one metre from the entry point and 1 metre apart from each other, (Figure 6.7 & 6.8) and allowed to investigate until the dog stopped investigating either of the objects for 10 seconds. The dog was then removed from the room by the owner calling the dog or going and taking hold of the dog. One of the objects was then changed from a bin to a constructed wooden block (which had also been cleaned with Safe4 disinfectant) (Figure 6.8), before the dog was brought back into the room to investigate again. The side of presentation of the novel object was counter balanced across the population of each test group to minimise side biases. After 10 seconds of no longer investigating either of the objects the dog was removed from the room. For a photograph of the test see Figure 6.9.

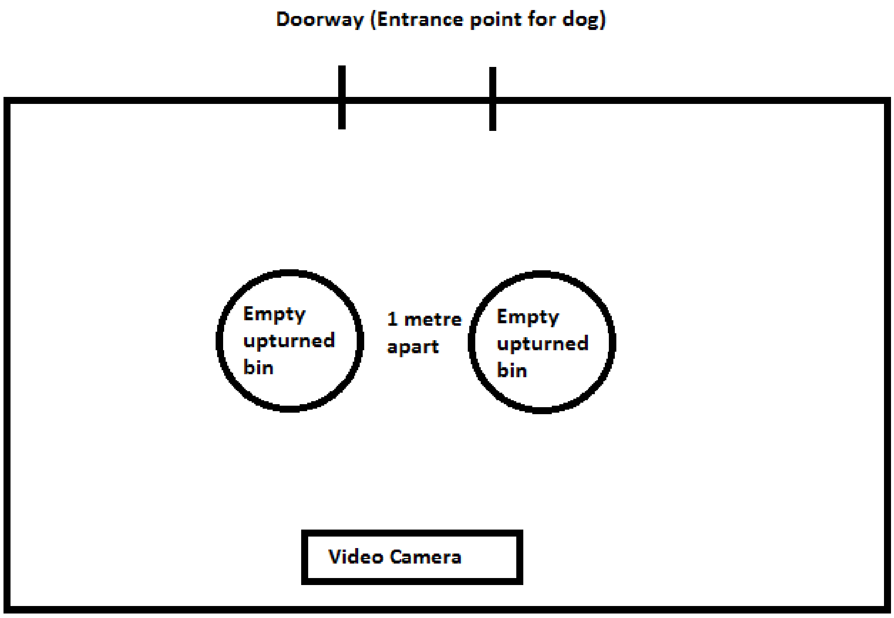


Figure 6.7 Diagram of the set up for the novel object investigation test on the dog's first exposure to the room where both objects were identical. Room measured 6m by 7m, the objects were 1 metre from the dog's entry point and one metre apart from each other.

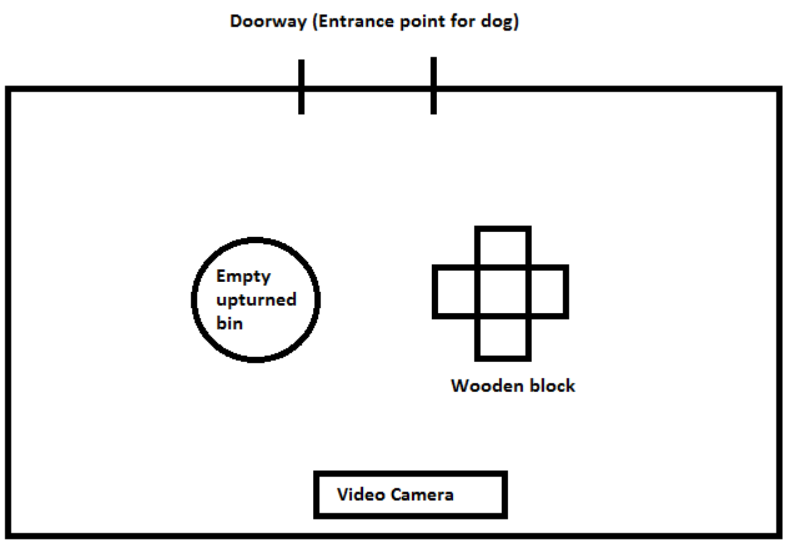


Figure 6.8. Diagram of the set up for the novel object investigation test on the dog's exposure to the room where one of the objects was replaced by a different object (wooden block). The objects

were 1 metre from the dog's entry point and one metre apart from each other. The side that the block was placed was counterbalanced across the study population of each group to minimise side bias.

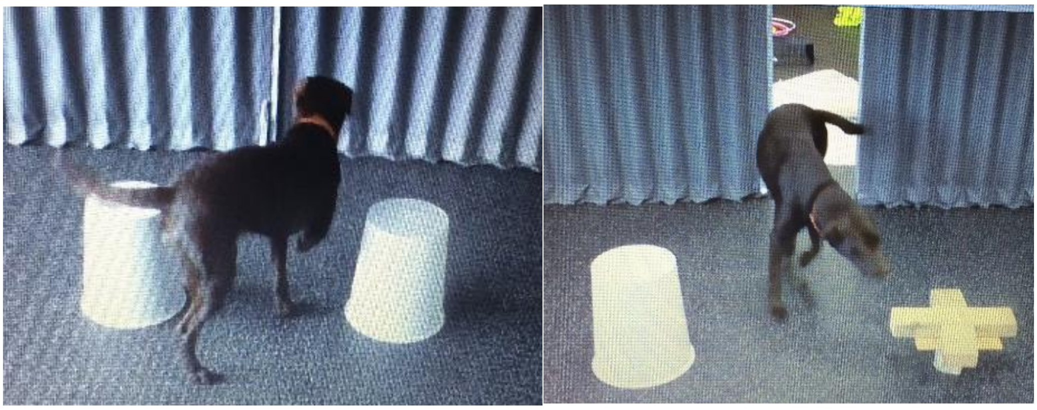


Figure 6.9 Photographs of the Novel Object test. The dog can be seen on the left in the room with the two matching objects, then on the right being sent into the room when one of the objects has changed.

It was hypothesised that dogs in the higher scoring group would spend more time investigating the novel object compared to those in the lower scoring group in this 2-part test. This is because they are predicted to be potentially more interested in new things, and so will more readily investigate change (DIAS elements 'my dog takes a long time to lose interest in new things', and 'my dog is very interested in new things and new places', PANAS 'your dog shows little interest in its surroundings' (reverse scored and therefore switch to the new routine more readily (DIAS element 'my dog reacts very quickly', and PANAS element 'your dog requires a great deal of encouragement to take part in energetic activities' (reverse scored)).

6.2.4 Attention Test:

Positioned seated in a chair directly in front of the dog, with the dog off lead and no-one else present in the room, the researcher showed the dog a treat (Akela 80:20 dog food). The treat was held in the fingertips of the researcher's hands (arms outstretched directly in front) and then the hands were moved apart from each other in a horizontal motion in front of the dog, the arms stopped moving when they were out to the researchers sides (See Figure 6.10). The treat remained visible to the dog during this process. The treat moved either with the right or left hand for multiple repetitions (10 in a row in the same direction before changing to the other direction until the dog followed the new direction immediately the hands moved for 10 trials in a row, or until the dog had experienced fifty trials without reaching the criterion (10 in a row after having changed to the opposite direction)). When the dog's gaze followed the hand containing the treat then the dog was given the treat. If the dog made a mistake then the time taken until the dog switched its gaze to the correct choice was recorded. If no switch of gaze was observed within 10 seconds or the dog walked away (defined as a mistake), both hands were placed behind the researcher's back, the trials were resumed after getting the dog's attention. If 3 refusals to participate (defined by the dog walking away or failing to make a choice for a block of 30 seconds) occurred in a row then trials were discontinued. See figure 6.10.

Figure 6.10 Photographs of the Attention Test. The dog can be seen following and choosing the hand with the biscuit in.

It was hypothesised that dogs in the higher scoring group may therefore complete the task in fewer trials compared to dogs in the lower scoring group. This is because they are predicted to be potentially more interested in new things, and so will notice the change in the pattern sooner (DIAS elements 'my dog takes a long time to lose interest in new things', and 'my dog is very interested in new things and new places', PANAS 'your dog shows little interest in its surroundings' (reverse scored and therefore switch to the new routine more readily (DIAS element 'my dog reacts very quickly', and PANAS element 'your dog requires a great deal of encouragement to take part in energetic activities' (reverse scored)). Dogs in the higher scoring group may also be likely to complete the task in fewer trials as their scores for "Energy & Interest" within the PANAS include the elements 'your dog is full of energy', and 'your dog is lazy' (reverse scored), meaning the dogs may be more likely to engage with the task.

6.2.5 Discrimination Test:

For this project, the dog was trained to nose target one cone as opposed to another, the cones differed by colour (one blue and one green, sides and colours randomised for each dog) and were presented 50cm in front of the dog, and 50cm apart from each other (for diagram see Figure 6.11, photograph 6.12). The dog was allowed to work off lead and the owner was in the room reading a book 15 metres away from the researcher and dog and asked to not interact with their dog.

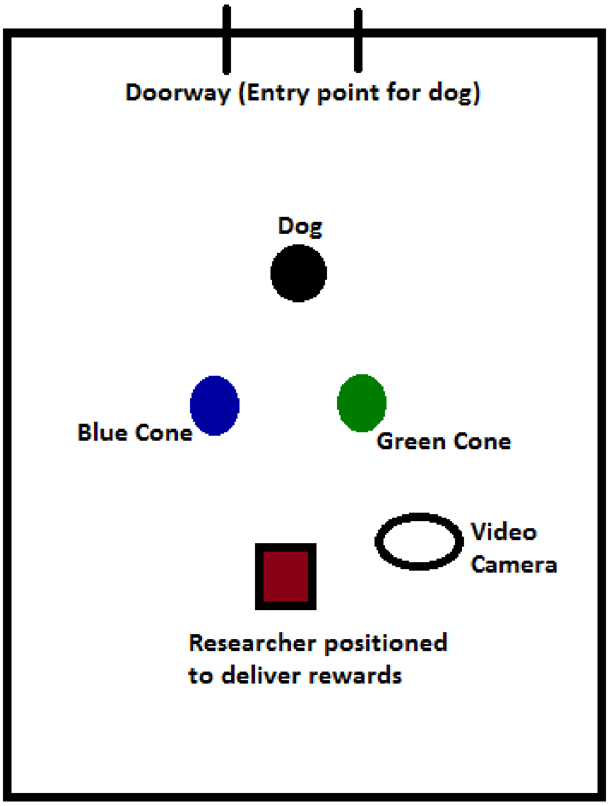


Figure 6.11. Set up for the discrimination test. The cones were 1 metre apart from each other and 50cmin front of the researcher, the dog start position was central to the cones but 50cm away, the room measured 6m by 7m.

Figure 6.12 Photograph of the Discrimination test. The dog can be seen targeting the cone they were trained to target with their nose.

For testing the researcher was seating in a chair and placed the cones in from of them. The cones were cleaned with Safe4 disinfectant before each dogs training. The dog was trained using food rewards (dog biscuits Akela 80:20, one biscuit delivered per correct response) for correct responses and ignored for incorrect responses. The cones remained in the same positions throughout training and testing but were picked up during breaks. Training occurred in blocks of 10 opportunities to perform the behaviour then the dog was given a break. For the first block of ten a treat was placed on top of the cone for the dog to eat. When the dog ate the treat off the cone the researcher said 'good' to mark the correct behaviour. The next block of ten started with 5 trials with a treat on top of the cone and then 5 trials where the treat was delivered away from the cone in response to the dog placing its nose on the cone to look for the treat. If the dog did not place its nose on the cone during these trials the previous training block was repeated. Correct responses were marked with the word 'good' before the treat was delivered. The next block of ten trials consisted of waiting for the dog to place its nose on the cone, marking the behaviour with the word 'good, and rewarding the dog away from the cone (treat placed on the floor in between the two cones) if the dog did perform the nose target. If the dog was unable to nose target the cone during this session the previous stage was repeated. If the dog was unable to meet criterion for the task after 100 repetitions (10 blocks of 10 trials) then training was stopped and the dog was recorded as unable to complete the task. When the dog reached criterion (8/10 correct responses in a row for 2 consecutive sessions with no treat on the cone) the dog was given a break and then given a 2 minute testing session in which they could offer the learned behaviour (nose target cone) as many times as they chose to. The dog was marked and rewarded with one biscuit for each correct response (biscuit placed between the two cones) and ignored for incorrect responses.

It was hypothesised that dogs in the higher scoring group (in terms of the DIAS scores for "Responsiveness" and PANAS scores for "Energy & Interest") would reach criterion in fewer trials based on the element within the DIAS "Responsiveness" element 'my dog is easy to train', and perform the learned behaviour more times during the testing period based on all of the elements within the PANAS relating to "Energy & Interest", compared to dogs in the lower scoring group.

6.2.6 Social Sensitivity Test:

The dog was given access to a room (6m by 3m) with an empty rubbish bin tipped over in the middle of the room (See Figure 6.13). The dog was shut in the room alone for ten seconds before a voice recording of the phrase 'what have you done? Bad dog' in a stern tone of voice by the researcher was played at 80 decibels. Following a further 1 minute along in the room after the recording the owner and researcher entered the room and then the owner took their dog for a10 minute break. The dog's reaction was filmed and the time spent interacting with the bin before and after the voice recording as well as the latency to interact with the bin following the sound of the voice recording was recorded. See Figure 6.13.

6.13 Photograph of the Social Sensitivity Test. The dog can be seen investigating the bin while alone in a room.

It was hypothesised that dogs in the higher scoring group (in terms of DIAS scores for "Responsiveness", the element 'my dog appears to be 'sorry' after it has done something wrong') would have a longer latency to investigate the bin following the voice recording compared to dogs in the lower scoring group as they are more sensitive to the social context.

6.2.7 Social Cues Test:

The dog was trained to approach two food bowls (1 metre apart and 3 metres away from the dog) following a signal of an arm pointing towards one of the bowls at any given time by the researcher who was positioned standing in between the two bowls facing the dog. One bowl was presented to the left of the dog, the other to the right of the dog. The dog was held by the owner who was seated on a chair opposite the researcher, when the pointing gesture was given by the researcher the owner released the dog to allow them to approach the bowls and make their choice. If the dog approached the correct bowl (the one the researcher was pointing at) the marker word 'good' preceded a food reward being given to the dog (dropped into the bowl the dog had approached). Training was conducted in blocks of 10 trials, within each block of ten trials the side which the point gesture was given was pseudorandomised so that there would be 5 points to each side in the session, but no more than 3 in a row pointing to the same side. Once the dog had made their choice and eaten the biscuit if correct, the owner was asked to neutrally call their dog back to them and take hold of them again. Once trained to discriminate between the bowls based on the signals (8/10 correct in a row over the course of one session), the dog was then given additional trials in blocks of ten where they were required to ignore the visual cues and go to the opposite bowl in order to obtain a reward. These tests were carried out in blocks of ten and pseudorandomised as before, if the dog did not reach criterion (8/10 correct in a session) over the course of five blocks of sessions the testing was discontinued. If the dog was incorrect on a trial, the bowls were picked up and the dog was moved away before trials resumed. If the dog refused to participate for 3 consecutive trials then the study was discontinued. See figure 6.14.

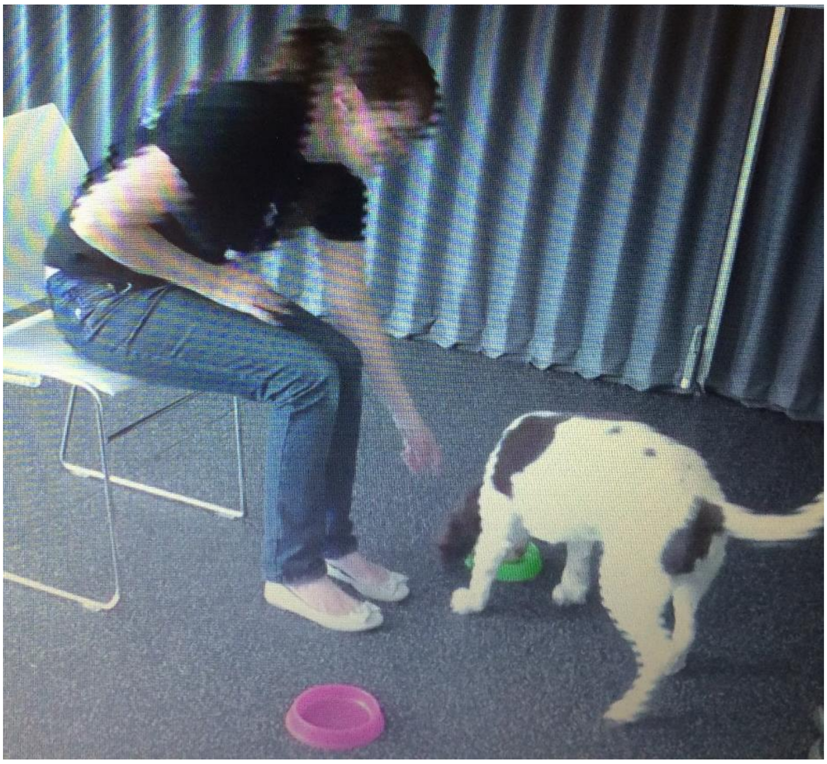


Figure 6.14 Photograph of the Social Cues Test. The dog can be seen selecting the bowl where the point cue was directed. The dog had the choice between two empty bowls, one of which would be pointed towards, if the dog followed the pointing gesture a biscuit would be dropped into the bowl.

It was hypothesised that dogs in the higher scoring group (in terms of the DIAS scores for "Responsiveness" and PANAS scores for "Energy & Interest") would reach criterion in fewer trials (based on the PANAS "Energy & Interest" element 'my dog is easy to train' and the DIAS elements 'your dog is full of energy', your dog is lazy' (reverse scored), and 'your dog requires a great deal of encouragement to take part in energetic activities' (reverse scored)) compared to the lower scoring group. Based on the element within the PANAS 'my dog appears to be 'sorry' after it has done something wrong' and 'my dog reacts very quickly', the higher scoring group were predicted to learn to ignore the social cues faster compared to dogs in the lower scoring group because they are more likely to respond and alter their behaviour when they get the task wrong.

6.2.8 Test Re-test

All dogs were retested 2-4 weeks after the initial test using the same process as during testing, in order to assess the reliability of the tests, however, it was anticipated that there may be differences in performance across test-retest overall due to learning effects or between the two groups based on their different temperament profiles. This complicates the interpretation of reliability. The implications of test-retest differences are discussed in relation to the specific findings of a given test.

6.2.9 Retest Hypotheses:

Based on the elements within “Responsiveness” about dogs scoring higher being easier to train, it is predicted that in the learning tasks (discrimination and attention tests) the lower scoring group may improve performance when re-acquiring the task at retest, while the higher scoring group may maintain their test performance.

6.2.10 Data Analysis

Video footage of the behaviour tests was analysed using Solomon Video Coding software and statistical analysis was performed using SPSS21.0. A proportion (10%) of the videos were blind analysed (allocated a random number linked to their original data, known ot the researcher but not to the student doing the blind analysis) by another individual (an undergraduate student in animal related studies from the university) to check for inter-rater reliability. Dogs which failed to complete the tests were not included in analysis for the tests which they failed to complete.

Shapiro Wilkes tests for normality were completed on all of the data and relevant parametric (independent samples t-test) or non-parametric (Mann Whitney test) tests were then used to examine significant differences between the two groups of dogs, and between the dogs at test and retest.

6.3 Results

Of the 21 subjects recruited for testing, 3 were unable to complete the discrimination test and 3 were unable to complete the attention test, in both cases 2 of the dogs were from the lower scoring group and 1 was from the higher scoring group, it was the same dogs withdrawn from each test. Their results for these tests were excluded but their results for the other tests were included in analysis.

For a full summary of all research findings from the study see Appendix 4.

6.3.1 Mapping Test results onto specific elements of PANAS and DIAS

Based on the design of the tests, certain test were proposed to relate to elements within "Responsiveness" from the DIAS and "Energy & Interest" from the PANAS. Results are broken down into these elements.

6.3.1.1 Novel Environment Test

The novel environment test was proposed to map onto the elements 'your dog shows little interest in its surroundings' (reverse scored), 'your dog is full of energy', 'your dog is lazy' (reverse scored), 'your dog requires a great deal of encouragement to take part in energetic activities' (reverse scored) from the "Energy & Interest" component of the PANAS. As well as the elements 'my dog takes a long time to lose interest in new things', and 'my dog is very interested in new things and new paces' from the "Responsiveness" component of the DIAS.

No significant differences were found between the test groups in the initial novel environment test. At retest the lower scoring group investigated significantly fewer objects compared to their performance in the test (Related Samples Wilcoxon Signed Rank, p=0.042), this was also reflected in the population analysis (Related Samples Wilcoxon Signed Rank, p=0.011). See Figure 6.15.

	Novel Environment Test		Higher Scoring Group test-retest analysis		Lower Scoring Group test-retest analysis		Population test-retest analysis	
	HS	LS	Test	Retest	Test	Retest	Test	Retest
number of objects investigated	8.6 +/-1.4	7.5 +/-2.2	8.64 +/-4.7	5.9 +/-5.5	7.5 +/-6.98	3.2 +/- 3.55	8.1 +/-5.77	4.56 +/-4.7
Test	Independent Mann Whitney U 0.426		Related Samples Wilcoxon Signed Rank 0.108		Related Samples Wilcoxon Signed Rank 0.042		Related Samples Wilcoxon Signed Rank 0.011	

Figure 6.15 Novel Environment Test Results comparing the number of objects investigated between the higher scoring (HS) group and the lower scoring (LS) group, and test retest comparisons for each group and population. Highlighted text indicates where a significant difference was observed with a p value less than 0.05.

6.3.1.2 Novel Object Test

The novel object test was proposed to map onto 'Your dog shows little interest in its surroundings' (reverse scored in questionnaire), 'your dog requires a great deal of encouragement to take part in energetic activities' (reverse scored in questionnaire) from the PANAS & 'My dog takes a long time to lose interest in new things' and 'my dog is very interested in new things and new places' from the DIAS.

In the novel object tests no significant differences were found between the two test groups in individual measures, but within the higher scoring group there was a significant difference in the sniffing time between sniffing the familiar object and the novel object at both test (Related Samples Wilcoxon signed rank, p=0.034) and retest (Related Samples Wilcoxon signed rank, p=0.05), which was not significant in the lower scoring group. See Figure 6.16.

Novel Object Test	Test		Retest	
	Higher Scoring Group	Lower Scoring Group	Higher Scoring Group	Lower Scoring Group
Block Sniffing Time (seconds)	1.4 +/- 0.4	1.72 +/- 0.7	2.42 +/-0.72	1.38 +/-0.8
Bin Sniffing Time (seconds)	0.15 +/-0.24	0.38 +/-0.24	0.46 +/-0.26	0.36 +/-0.13
Related Samples Wilcoxon signed rank (p)	0.034	0.14	0.05	0.225

Figure 6.16. Results for Novel Object Test comparing the length of time spent sniffing (+/- standard deviation) the familiar object (the bin) and the novel object (the block) for both the higher scoring (HS) and lower scoring (LS) groups at test and retest. Highlighted text indicates where a significant difference was observed with a p value less than 0.05.

6.3.1.3 Attention Test

The attention test mapped onto the elements 'Your dog shows little interest in its surroundings' (reverse scored in questionnaire), 'your dog requires a great deal of encouragement to take part in energetic activities'(reverse scored in questionnaire) from the PANAS & 'My dog takes a long time to lose interest in new things' and 'my dog is very interested in new things and new places', and 'my dog reacts very quickly' from the "Energy & Interest" component of the DIAS.

There were no significant differences in the number of trials to complete the attention test, although within the higher scoring group there was a significant decrease in the number of trials to complete the task between test and retest (Related Samples Wilcoxon Signed Rank, p=0.042) and as a population a significant decrease in the number of times the dogs opted out from test to retest (Related Samples Wilcoxon Signed Rank, p=0.036), but no significant differences were observed within each group. See Figure 6.17.

	Attention Test		Higher Scoring Group test-retest analysis		Lower Scoring Group test-retest analysis		Population test-retest analysis		Population test-retest analysis		
	HS	LS.	Test	Retest	Test	Retest	Test	Retest		Test	Retest
Trials to complete	28.1 +/- 2.1	31.8 +/- 5	28.1 +/- 6.6	24.2 +/-6	31.75 +/-14.05	27.2 +/-9.11	29.72 +/- 10.38	25.79 +/- 7.72	Opt out	5.4 +/-4.71	2.31 +/-2.52
Test	One Way Anova F = 0.535, p=0.475		Related Samples Wilcoxon Signed Rank .p=0.042		Related Samples Wilcoxon Signed Rank .p=0.123		Related Samples Wilcoxon Signed Rank 0.014		.	Related Samples Wilcoxon Signed Rank 0.036	

Figure 6.17. Results for Attention Test comparing the higher scoring (HS) and lower scoring (LS) groups, performance in terms of number of trials to complete the task, and comparing test retest performance in the higher scoring group, lower scoring group, and the study population in terms of trials to complete the task. Population statistics for opting out of the attention task are also featured. Highlighted text indicates where a significant difference was observed with a p value less than 0.05.

6.3.1.4 Discrimination Test

The discrimination test mapped onto the elements ‘your dog is full of energy’, ‘your dog is lazy’ (reverse scored), ‘your dog requires a great deal of encouragement to take part in energetic activities’ (reverse scored), from the responsiveness component of the PANAS, and the element ‘my dog is easy to train’, from the “Energy & Interest” component of the DIAS.

There was a significant difference between the higher scoring and lower scoring groups in terms of the number of trials taken to learn the discrimination task at test (Independent Sample Mann Whitney U Test, p=0.021) but not retest (Independent Sample Mann Whitney U Test, 0.46), with the higher scoring group taking fewer trials initially. The lower scoring group learned the discrimination task in significantly fewer trials in the retest (average 31.8 trials, and more in line with the higher scoring group’s performance) compared to their performance in the test situation (average 88.6 trials, Related Samples Wilcoxon Signed Rank, p=0.009). The higher scoring group showed no significant difference in the number of trials taken to learn the discrimination task between test and retest (Related Samples Wilcoxon Signed Rank, p=0.553). See Figure 6.18 below for breakdown of results.

	Test Data		Retest Data		Population Analysis		higher scoring Group		lower scoring Group	
	HS	**LS**	**HS**	**LS**	**Test**	**Retest**	**Test**	**Retest**	**Test**	**Retest**
Trials to learn	46.5 +/-8.9	100.7 +/-14.7	28.3 +/-2.2	31.8 +/-2.87	70.41 +/-42.16	32.82 +/-9.74	44.4 +/-26.6	34.32 +/-11.2	88.6 +/-42.37	31.8 +/-9.08
Test Results	Independent Sample Mann Whitney U Test **p=0.021**		Independent Sample Mann Whitney U Test .p=0.46		**Wilcoxon Signed Rank test .p=0.007**		**Related Samples Wilcoxon Signed Rank** .p=0.553		**Related Samples Wilcoxon Signed Rank .p=0.009**	

Figure 6.18. Results for the Discrimination Test comparing performance between the higher scoring (HS) group and lower scoring (LS) group in terms of the number of trials taken to reach criterion. Highlighted test indicates where a significant difference was observed with a p value less than 0.05.

6.3.1.5 Social Sensitivity Test & Social Cues Test

The social sensitivity test was mapped onto the 'my dog appears to be 'sorry' after it has done something wrong' element from the "Energy & Interest" component of the DIAS. The social cues test was mapped onto the elements 'your dog is full of energy', 'your dog is lazy' (reverse scored), 'your dog requires a great deal of encouragement to take part in energetic activities' (reverse scored), from the responsiveness component of the PANAS, and the elements 'my dog appears to be 'sorry' after it has done something wrong', 'my dog is easy to train', and 'my dog reacts very quickly', from the "Energy & Interest" component of the DIAS. No significant effects were observed for either of these tests in relation to the two populations. For full summary of results see Appendix 4.

6.4 Discussion

6.4.1 Overview

While a number of tests and parameters were examined, only one test result revealed a significant difference between the high and low scoring groups at test, and this was not observed at re-test. While other interesting differences were found these need to be interpreted cautiously as reliability was not observed across time periods for the tests. There is also the issue of multiple testing and the number of measures taken meaning a high probability of finding a significant result by chance. Individual test results are discussed in detail in the following text.

6.4.2 Social Tests

The hypothesis for the social cues test that "dogs in the higher scoring group would have a longer latency to investigating the bin following the voice recording compared to dogs in the lower scoring group" was rejected. As was the hypothesis for the social sensitivity test that "dogs in the higher scoring group would reach criterion in fewer trials and the higher scoring group were predicted to learn to ignore the social cues faster compared to dogs in the lower scoring group." No significant results were identified in either study.

Test design may not have conducive for measuring these elements. The dogs' learned history of social cues (for example a learned history of following human pointing cues) may have impacted upon their performance, as found in Lazarowski *et al* 2015, where research dogs which had limited social contact with humans failed to perform as well as pet dogs on a distal pointing task. As pet dogs appear to be more tuned in to human cues, they may have found it harder to ignore the pointing gesture in reversal learning (Milgram *et al*, 2004). Potentially a reversal task within human social cues may have elicited more accurate results, but as a number of the questions from the PANAS and DIAS making up the elements of interest had social components this was not initially considered to have been likely to cause an issue. It may have been the case that the dogs may not have learned the cue, the behaviour may have been a result of stimulus enhancement (McKinley & Sambrook, 2000), and insufficient time to learn an alternative behaviour pattern may have been allowed. Attentiveness to human cues is thought to have an evolutionary origin (Scott & Fuller, 1965) as the domestication process necessitated dogs to attend to and work with humans. Dogs follow human pointing and look to their owner in situations where they are problem solving, this could represent responsiveness to people and in tasks such as the social cues task, and potentially measuring attempts to communicate with the owner may provide more information (Miklosi *et al*, 2003). These factors may explain why all of the dogs struggled to ignore the human social cue and try an alternative solution. The social tests are likely impacted on by past learning, especially the social sensitivity test as although some dogs may naturally be more sensitive to the stern voice, sudden sound, or punishment which could reflect on temperament it is likely that a history of being told off, or situations involving stealing are impacting on the dogs response to this test (Rooney & Cowan, 2011).

6.4.3 Novel Environment Test

The hypothesis that "dogs in the higher scoring group would spend significantly longer investigating than dogs in the lower scoring group" was rejected.

The novel environment test found that dogs in the lower scoring group investigated significantly less on the retest exposure compared to their performance in test. No differences were observed in the higher scoring group. This is potentially linked to the DIAS Responsiveness element 'my dog is very interested in new things and new places', as dogs in the lower scoring group would be expected to lose interest in the novel environment more quickly so are less likely to investigate a second exposure. There was no observable differences between the groups on their first exposure which may potentially relate to a lack of difference between the tests populations in terms of their scores for the PANAS element ‘your dog shows little interest in its surroundings’, or the test was not sensitive enough to tease out any difference. Differences in the PANAS elements ‘your dog is full of energy’, and ‘your dog is lazy’ do not appear to have been revealed by this test, had they been it would have been expected to have seen differences in the groups in terms of time spent investigating, with those in the higher scoring group spending more time investigating at both test and retest. Gabora & Colgan, 1990 found that in rats exposed to a novel environment exploratory behaviour was inhibited, this could account for the lack of significant results. Potentially the dogs needed habituation to the room before the testing session, although this would take away the novelty element of the test it may have provided more useful information on exploratory behaviour.

6.4.4 Novel Object Test

The hypothesis that “dogs in the higher scoring group would spend more time investigating the novel object compared to those in the lower scoring group” was rejected.

Dogs in the higher scoring group spent longer investigating a novel object compared to a familiar object, an effect that was not seen in the lower scoring group. This is consistent with higher levels of the responsiveness element ‘my dog is very interested in new things and new places’, as dogs in the higher scoring group appear to show more interest in their environment, which could explain their increased level of interest when something changed. In relation to the previous study on working dog data this could reflect the dogs in the withdrawn group showing less interest in their environment or lose interest quickly, which could impact on their ability to work as it may affect how focused they remain on their task. Those interested in their environment are likely to search more effectively and investigate more thoroughly which has been indicated in desirable traits of specialist search dogs (Rooney, 2004).

6.4.5 Attention Test

The hypothesis that “dogs in the higher scoring higher scoring group would complete the task in fewer trials compared to dogs in the lower scoring group” was rejected.

This test was unable to differentiate between the high and low scoring populations. This test may not have been well suited to measuring attention, but some of the results could indicate that the rates of learning between the two groups differed. Dogs in the higher scoring group completed the attention task in significantly fewer trials in the retest compared to the test. This was thought to be linked to the element ‘my dog reacts very quickly’, as the higher scoring group would likely react faster to a change in pattern and therefore complete the task in fewer trials compared to those in the lower scoring group. The PANAS elements ‘your dog is full of energy’, and ‘your dog is lazy’, may also be revealed in this test as they could link to task engagement, so higher scoring individuals are likely to readily engage with the task and complete it in fewer trials. It is possible that the impact of these elements was revealed here but not in the novelty tests because the latter tests did not involve social engagement which is potentially important to these elements, i.e. A training aspect is required to assess them. The attention test results, in terms of the dogs in the higher scoring group completing the task in fewer trials, also highlight that the element 'my dog reacts quickly' is one of potential importance in terms of being relevant in the differences between the groups. It could be that this test was more directed towards measuring rates of learning as the higher scoring group seemed to have improved performance in terms of number of trials to complete the task on the second exposure, while the lower scoring group remained similar at retest to test. While this could be due to the temperament profiles of the dogs in the higher scoring group, it could also be a result of other factors such as previous training history which can impact of learning (Rooney & Cowan, 2011). As the test results changed on retest for the higher scoring group this suggests that the test is unreliable, although for the lower scoring group performance was unchanged at retest, so while reliability is being questioned it could be a result of differing rates of task acquisition in the different groups of dogs.

6.4.6 Discrimination Test

The hypothesis that "dogs in the higher scoring group would perform the learned behaviour more times during the testing period compared to the lower scoring group" was rejected.

The hypothesis that "dogs in the higher scoring group would reach criterion in fewer trials" compared to the lower scoring group" was accepted.

Dogs in the higher scoring group acquired the discrimination task in fewer trials compared to the lower scoring group. This may reflect their higher levels of trait responsiveness "Energy & Interest", making them easier to engage with and quicker to learn a new task. This could be the result of elements being measured within the responsiveness construct within the PANAS 'your dog is full of energy' and the reverse scored item 'your dog is lazy', as an element of the level of engagement the dogs showed with the task. However, at retest no differences were seen between task acquisitions between the two groups, this could suggest that while initial task acquisition takes longer for dogs in the lower scoring group, once the task is acquired they retain that information in a similar way to dogs in the higher scoring group. This explanation seems more likely than inherent unreliability of the test as the results in the retest are quite similar between the two groups indicating a common level of performance with experience. This test potentially differentiates the two groups in relation to the element 'my dog is easy to train' of responsiveness within the DIAS, as dogs in the higher scoring group initially learned the task in fewer trials they could be described as more ‘easy to train’. The results observed in this test could be related to previous training history and experience. Dogs with a history of lots of training and dogs that have been trained using positive reinforcement acquire a new task faster, as reported in Rooney & Cowan, 2011 in a similar discrimination task to that used within this project that was used to examine the impact of type of training on task acquisition. It may have been the case that in the present study the dogs in the high scoring group had a greater history of positive reinforcement training, as this wasn’t controlled for in the selection of the dogs for the study it is an unknown factor, although the methods used to train all of the dogs within the study were the same.

6.4.7 Elements of interest

There is potentially a divide between results associated with tests measuring elements related to the physical environment and tests related to measuring engagement with people or tasks. It is possible that within these elements finer sub elements are responsible for the differences observed. For instance, ‘your dog is full of energy’ and ‘your dog is lazy’ (reverse scored), may consist of sub elements relating to engagement with social factors (such as ‘your dog is full of energy when working with you’) and physical environmental factors (such as ‘your dog is full of energy when exploring it’s environment’), which may not necessarily correlate with each other. The dogs recruited for the study may have been scored higher on one sub-element relative to another; this could explain why the tests aimed at assessing these elements were not all successful. The tests were designed based on assumptions that they would measure certain elements, as a number of the tests failed they were likely not well designed to measure what they were thought to be measuring, or other factors were responsible for the test results that extended beyond the elements they were thought to measure.

6.4.8 Other Factors

Factors other than temperament may also influence performance outcomes. Trainability has been report to be affected by breed (Turcsan et al, 2011) which may or may not relate to temperament or cognitive abilities, but research on sex and neuter status has shown mixed results in terms of their impact on trainability in dogs. While some studies have found no effect in dogs (Serpell & Hsu, 2005. Turcsan et al, 2011) and horses (Gorecka-Bruzda & Jezierski, 2010), endocrinological effects including those related to gender have been shown to impact components of memory, which may impact novel object test and task learning (Sutcliffe *et al*, 2007), which would impact on trainability. This could link in to the elements, ‘my dog is very interested in new things and new places’, and ‘my dog is easy to train’ and may have had an impact on the novel environment test, the novel object test, and the discrimination test, although a larger study population would be required to test the effects of gender or breed.

Highly trained dogs have been shown to spend longer interacting with tasks and objects, while untrained dogs have been shown to spend more time looking back to their owner during problem solving tasks, suggesting that level of training history may improve problem solving ability (McKinley & Sambrook, 2000. Marshall-Pescini *et al*, 2008). Training has also been shown to have an effect on the response of horses in a novel object task (Visser *et al*, 2002) and past behaviour can influence future behaviour in both new and familiar situations (Danner *et al*, 2008). Based on this evidence previous training experience may have impacted on performance in the discrimination test because of the training element within it, the attention test because of the problem solving element, and the social cues task as it involved problem solving. So as well as temperament, certain learned components of behaviour may impact on field performance, and learning over time in the field may result in behaviours being acquired that are not constructive to work and may result in withdrawal from service.

6.4.9 Limitations

Within behaviour tests there is always the issue of having to rely on measurable outcomes to determine results, but behaviour is a complex notion to understand. Tests are designed to measure behaviours and traits but how well they can measure these, or if they are measuring what they were designed to measure is often difficult to determine.

The issues with reliability highlighted in this study show how the effects of learning can impact on repeated trials. So while for example the discrimination test was not shown to have retest reliability, it did highlight differences in rates of learning. So just because a test is not reliable doesn't mean it is not useful, and this has highlighted an issue that has been faced previously in the working dog literature (see Chapter 3). This creates an issue as if performance is the same at retest as test it is therefore reliable, while a difference between test and retest means the test is unreliable. However, the test still provides information, which can still be used to help understand behaviour so shouldn't necessarily be dismissed, just interpreted with caution.

The sample size within this study was small because dogs had to be recruited to match a specific temperament profile from the pet dog population. Had a larger sample been possible to recruit in the time frame more conclusive results may have been produced. There is also the issue that the study was conducted on a pet dog population, rather than a working dog population due to recruitment restrictions, working dogs may perform differently in the tests because of their prior levels of training and it was identified in Chapter 5 that working dogs generally differ from the pet dog population. It may have been that the sample was difficult to recruit because dogs were being selected for traits that appear in working dogs, and these are not the norm in a pet dog population.

There is still a need to define population norms in working dogs for the tests to establish reference values. This would require testing a significant number of working dogs scoring both within the higher scoring and lower scoring ranges to identify reference values and also potentially dividing these into working roles as well as training methods used, as identified these can impact on performance in training tests and investigatory behaviour (Rooney & Cowan, 2011).

The results of the previous study in Chapter 5 identifying temperament traits associated with withdrawal in working dogs have helped to identify factors associated with working dogs in general, although different working dog roles may require the selection of different traits, or different levels of the same traits (Rooney *et al*, 2004). There is a need to objectively define success in working dogs in a variety of disciplines and then develop behaviour tests to measure such traits for selection purposes (Wilsson & Sinn, 2012).

6.5 Conclusions

This is new information in terms of developing behaviour tests to identify and correlate with these traits from the PANAS and DIAS. Dogs falling into the higher scoring or lower scoring group (in terms of components from the PANAS and the DIAS) could be differentiated based on performance in the discrimination test based on the number of trials taken to acquire the task. Further information is still needed in terms of how working dogs perform on the tests and what cut off points would be obtained for separating dogs into high and low scoring groups based on their test performance. There is also a need to establish if the temperament profiles observed in withdrawn dogs are a result of them failing in their work, or if the profile is responsible for their poor performance leading to withdrawal before this could be put into use in terms of an additional tool for working dog selection. Further work within the working dog sector is necessary to make the tests practically useful in terms of the selection process and identifying individuals at risk.

Chapter 7: Discussion

7.1 Introduction:

Current research has made solid contributions to working dog selection processes and aided understanding that behavioural characteristics are often responsible for failure rather than physical or sensory attributes (Foyer *et al*, 2014. Maejima *et al*, 2007. Rooney *et al*, 2007). Working dog selection is an area of ongoing research where methods of assessment of working dogs are constantly being developed and revised to prevent withdrawal in the field due to behaviour (Sinn *et al*, 2010). Generally traits such as confidence and desire for work have been shown to be of importance in terms of working dog certification (e.g. Maejima *et al*, 2007), although working dog withdrawal from service is still an issue. This research has sought to work towards improving working dog selection through understanding the relationship between temperament traits measured using the PANAS and DIAS and working dog withdrawal from service. Specifically it has aimed to develop methodologies based around temperament traits to facilitate the pre-selection of working dogs and identify individuals at risk of withdrawal from service. Existing research has merits in identifying fear related behaviours that contribute to withdraw from service in working dogs, although other factors of temperament related to resilience and distractibility likely have merit in terms of identifying working dogs at risk of withdrawal, as outlined in this project.

Despite processes, such as selection and certification tests being in place to ensure working dogs have successful field careers, a proportion of dogs that are selected for service fail to qualify (Foyer *et al,* 2015), and a proportion of those that qualify end up being withdrawn from service due to behavioural problems (Evans *et al*, 2007). Not only does withdrawal from service have welfare implications for the dogs if they have proved unable to cope in a working environment, it also has financial implications in terms of initial and ongoing investment of time and resources in dogs that prove to be unsuitable. The other major concern with dogs being withdrawn from service due to a failure to perform in the field, is that there has already been potential risk to the handler and public safety. It is therefore imperative to develop measures to assess working dogs as early as possible to determine which will be able to perform effectively for the entirety of their working life.

7.2 Arousal and Performance

Arousal levels may be of importance in terms of performance at either task or individual level as per the Yerkes Dodson (Yerkes & Dodson, 1908) and Individual Zone of Optimal Functioning (IZOF, Hagvet & Hanin, 2007) laws. In working dogs, withdrawal from field service when dogs have passed selection and training may be a result of the pressures of the working environment, that are not able to be replicated in training, impacting on arousal levels either in an excitatory or inhibitory manner. Information on how arousal impacts performance could result in the development of mitigating measures to ensure optimal performance, through manipulating arousal to ensure dogs are kept in their 'optimal' range.

It is likely that individual dogs are differentially aroused by the same stimuli, making it difficult to ascertain level of arousal being elicited by certain tasks or situations without a physiological measure (Paul *et al*, 2005). This study was limited by its use of assumed arousal, as opposed to having a physiological measurement to confirm it. It may be useful to assess coping style and performance, rather than arousal as there is evidence to support the idea that learning is impacted on by coping styles. For example, dogs respond to stressful and distracting situations differently (Horvath *et al*, 2007), and this is thought to be related to individual differences in coping styles (Koolhaas *et al*, 1999). Temperament is thought to affect coping strategies, with passive copers showing relatively consistent strategies both within and across a number of contexts, while active copers vary their strategies in both similar and different contexts (Seaman *et al*, 2002). Blackwell *et al*, 2010, found that kennelled dogs experiencing stress adopted either a reactive coping strategy (showed fearful behaviours and/or reduced learning) or a proactive coping strategy (showing increased learning ability of new tasks). This is important in terms of predicting working dog performance in a variety of contexts as those with more consistent and proactive strategies are likely to be more predictable across contexts and potentially more reliable in their responses. If they have favourable coping strategies conducive to a working environment this could be useful in predicting success in working dogs. Dogs with less consistent strategies are likely to be less predictable when operating under stressful conditions which is potentially dangerous in a working setting. Predictable dogs with consistent strategies can be managed more effectively, making them more suitable for work and less of a risk as their responses and behaviour are more likely to be anticipated allowing for effective management strategies to be put in place by handlers. It could also be the case that passive copers with more consistent strategies are more likely to behave similarly in selection and certification tests to in an operational context.

The results obtained within this part of the research demonstrated that it is more likely that optimal arousal for performance is more individual based rather than task based although that is not to the exclusion of task arousal theories, but further investigation utilising more refined description and quantification of arousal would be valuable to support or refute this.

In light of this there was not a clear path to follow in terms of investigating the relationships between arousal and field performance based on the data and existing methodologies. Although temperament was indicated as having a potential effect on performance and this has been observed in the literature previously (Seaman *et al*, 2002). This resulted in the consideration of aspects of working dog temperament which may stretch beyond arousal and could inform regarding which dogs are most suitable for working tasks. To establish this there was a need to examine how dogs are being selected for working roles, and what constitutes a successful working dog. The aim was to try to determine where potential gaps in the selection processes lie as it appears that individuals who later prove to be unsuitable for work are passing the selection process.

7.3 Selection Tests

While selection tests have been shown to be able to identify temperament traits related to passing certification (Asher *et al*, 2013), there are problems with such tests which were highlighted in the systematic review (Chapter 3) which revealed that while current selection assessments provide useful tools for helping to select dogs for working roles, there is still room for improvement and a need to understand how the temperament qualities they found to be of relevance in terms of passing certification actually appear in working dogs in the field (Wilsson & Sinn, 2012), successful or withdrawn. The review also identified the lack of consistent terminology in describing temperament traits and the problems this can give rise to in particular the use of the shyness (Svartberg, 2002) or fearfulness (De Meester *et al*, 2011) and confidence (De Meester *et al*, 2011) or boldness (Svartberg, 2002). Another issue is the use of continuums as shyness/boldness (Svartberg, 2002) or confidence/fearfulness (De Meester *et al*, 2011) is that confidence/boldness are different to fearfulness/shyness and should be measured independently. For example a dog can be both confident around people but fearful of noises, in creating one score based on these elements the issues the dog has could be lost in the score whereas examining them independently could enable remedial training to minimise the issues relating to factors such as noise sensitivity to the point where the dogs could be suitable for work.

The lack of consistent use of behaviour terminology used in selection tests, makes it difficult to objectively compare measurements. In Chapter 3 an attempt was made to look across tests in order to identify where similar traits might be being measured directly or indirectly as part of a more complex trait, for example the use of the gunshot test to assess noise sensitivity on a broader trait level is examining a component of sensitivity to aversives.

The review further highlights the need to establish how such traits appear in both successful and withdrawn individuals. It is important to follow up on the dog's status and when certified whether they continue in service or end up being withdrawn, to identify how accurate the tests are at predicting lifetime service. This will improve understanding relating to the sensitivity and specificity of the tests. Then there is also the issue of individuals that fail selection but could have gone on to be successful working dogs, without following such dogs up with research it will never be certain if they truly were unsuitable for service.

Furthermore, selection needs to be based on meeting criteria. Rejection or qualification at certification in guide and service dogs (Duffy & Serpell, 2012), may be based on those that were the best out of a group being assessed, rather than those meeting specific thresholds. This is because often the number of dogs to choose from is limited, and there is a need to select a certain number of dogs per cohort. This can result in large variation in the quality of the working dogs being selected, as one cohort may consist of largely unsuitable dogs, but some must be selected regardless, possibly resulting in them failing certification tests or being withdrawn from service further down the line.

The review also identified the need for a centralised point of reference for working dog selection, testing and training for use by all those involved in these processes to ensure the same standards and procedures are being adhered to and that information on testing is being shared effectively. This should be based on an understanding of how established traits appear in successful and withdrawn working dogs, rather than based on the idea of what makes a good working dog. The assessment of temperament traits is well illustrated in pet dogs by the exploration of the trait "impulsivity" by Wright and others (Wright *et al* 2011 & 2012, Reimer *et al*, 2014). Using this approach they have not only been able to describe the trait psychometrically, but found laboratory and physiological correlates, evidence of the stability of the trait (Reimer *et al*, 2014) and, most recently, identified potential genetic correlates (Fadel *et al*, 2016). This project looked to further these ideas and establish if they would be of relevance to working dog selection.

7.4 Methods for assessing temperament

Impulsivity concerns the extent to which an individual evaluates the consequences of their actions, and relates to how quickly an individual reacts in a range of contexts (Wright *et al*, 2011). This makes it potentially an important overarching trait for working dogs. With the existing research into trait impulsivity in dogs as well as the likelihood of the trait influencing a range of aspects (such as trainability, how quickly a dog reacts, and interest in the environment) pertaining to performance in dogs, it was logical to focus on this trait as important in the assessment of working dogs as previously it has not been examined in this context. Existing measures for assessing impulsivity in pet dogs include questionnaire measures (the DIAS, Wright *et al*, 2011)), physiological measures in the form of urinary metabolites of serotonin and dopamine (Wright *et al*, 2012), as well as behavioural testing using a delayed reward paradigm (Wright *et al*, 2012). While physiological measures can be used to assess the trait analysis is costly in both time and monetarily, and like the behaviour test the results can only give an overall view of the trait, not the finer elements within.

While robust, the questionnaire is not useful at the selection stage alone as it requires a consistent caregiver to reliably answer the questions for the dog in a range of contexts, and the physiological and behavioural measures are time consuming making them potentially impractical for assessing working dogs. There was therefore a need to simplify methods for assessing impulsivity in dogs to create a short, practical test. Based on the use of tolerance to incremental delay used in the delayed reward paradigm, delay tolerance appeared to be a useful behavioural measurement system of DIAS scores for impulsivity. The nature of temporal delay tasks means that they involve lengthy training, so the decision was taken to attempt to use incremental spatial delay to attempt to assess impulsivity in dogs. This led to the development and refinement of a Spatial Discounting Test to enable fast and practical assessment of trait impulsivity in a field setting.

In addition it was found that the behavioural measure of the spatial discounting task could be simplified for the field. The simplified field study with reduced controls showed convergent validity with the laboratory study. It revealed that visual barriers and a standardised study environment could be excluded and reliable results still obtained. The research highlighted the ability to simplify tests for field use without significantly compromising predictive validity. However, in young dogs aged 2-9 months the spatial discounting task was not found to be a useful tool to assess impulsivity as their performance differed at re-test but the questionnaire measure was stable so this may be a useful tool to measure the trait in young dogs. The trait may be stable across ages because the owner's opinion is fixed early on about their dog so respond similarly on second completion of the questionnaire, the questionnaire has been shown to be stable up to 6 years after the initial assessment (Reimer *et al*, 2016). This is potentially due to the test not being suitable for younger dogs, or the trait not being stable in young dogs. If the trait is unstable in young dogs or still developing it may indicate that there is room to shape a young dog towards a desired temperament with appropriate training. This would be an important application in working dogs as it could enable borderline scoring dogs to be shaped to improve their chances of success in the field. Further research would be needed to establish how the trait develops with age and if it is susceptible to interference through training.

Using a spatial paradigm to test delay tolerance in dogs was a risky decision, as previous research has found that temporal and spatial delay were responded to differently in different species of primates (Stevens *et al*, 2005). This was a calculated risk as the measurement used was an incremental delay system to assess delay discounting, rather than a preference for temporal or spatial delay. The use of incremental delay may also be why the spatial and temporal tests showed similar results in terms of increased delay tolerance correlating with lower levels of impulsivity, it may be that in preference tests of delay the dogs do show greater tolerance of one type of delay compared to the other as observed in primates (Stevens *et al*, 2005).

7.5 Assessing working dog withdrawal in terms of temperament traits impulsivity and core affect.

Distraction either by positive or negative events is anecdotally a concern in regard to working dog withdrawal, and this can relate to a combination of impulsivity with sensitivity to positive and negative events (core affect). Trait distractibility has been found to be of importance in terms of working dog performance (Maejima *et al*, 2007), so examining these likely linked underlying mechanisms using validated tools could aid selection processes. Fortunately there is an instrument available for scaling this predisposition in dogs (PANAS scale). This along with the DIAS as a measure of impulsivity, could be used to assess working dog subjects on the basis of these traits.

With a view to assessing working dogs using PANAS and DIAS to identify key components related to withdrawal, an online survey was distributed using a variety of channels to collect data relating to core affect, impulsivity and demographic information on police and military working dogs in active service and those withdrawn from service.

Analysing a working police dog population in terms of PANAS and DIAS as measures of core affect and impulsivity respectively revealed components of the reported temperament traits correlated with withdrawal from service. However, while the reported traits are linked to withdrawal from service it is unclear whether the reported traits result in withdrawal, or whether failing in service results in changes to the reported traits. The components 'Responsiveness' from the DIAS, and "Energy & Interest" from the PANAS identified as relevance in terms of working dog withdrawal might affect performance because they relate to how engaged the dog is with their handler and the environment. The other components measured by the PANAS and DIAS that did not appear to be linked to withdrawal may still impact upon performance but the differences in them between successful and withdrawn dogs may be too subtle to notice. This is why handler scores were examined to try to tease out these factors, although a lack of differentiation between working roles in the dogs in the study may also account for this, but due to the sample size in this study, breaking down the population into their working roles to look for differences resulting in withdrawal would have resulted in very small sample sizes, limiting the conclusions that could be drawn from any results even further.

A limitation of this research is that the military dog population only consisted of dogs that were in active service as there appeared to be difficulties in gaining access to populations of withdrawn military working dogs. However, the military working dogs in active service closely resembled those in the police dog population in active service suggesting that the reported traits apply across these services. Further research into withdrawn military working dogs would be necessary to determine how the reported traits appear in the population. There is also a need to profile the dogs from a young age to determine how the reported traits develop and / or change and whether or not they can be used to predict successful or withdrawn outcomes in working dogs.

Comparisons were also made between the working dog population and the pet dog population (taken from Wright *et al*, 2011) to determine if certain elements of impulsivity and core affect were being selected for in working dogs. The police dog population scored higher than the pet dog population in terms of impulsivity components "Behavioural Regulation" and "Responsiveness", suggesting a positive selection for these reported traits. This seems logical as these elements include an interest in the environment as well as being easy to train. For working dogs the quicker they can be trained the easier it is to get them into work and provide remedial training if their performance starts to drop. An interest in the environment is also likely important as it indicates attentiveness stimuli such as visual and olfactory which are important in a number of working dog roles, although too much interest in the environment could contribute to distraction which has been shown to be related to withdrawal from service (Maejima *et al*, 2007). The police dog population scored lower for core affect component "Negative Activation", which suggests that working dogs are being selected to be less sensitive to punishing or negative stimuli, which links in to a need to have resilient working dogs that don't exhibit fear behaviours (Murphy, 1995. Ruefenacht *et al*, 2002. Batt et al, 2008. Gosling et al, 2009. Caron-Lormier *et al*, 2016. Rooney *et al*, 2016. Ilska *et al*, 2017). Working dogs often encounter potentially stressful environments, and being less sensitive to negative stressors would make them quicker to recover from stressful exposures and be less impacted by such stressors in general, enabling them to work more effectively in challenging conditions compared to dogs that are more sensitive to negative qualities within their environment. The police dog population scored higher for "Energy & Interest" as well, which again could be because they are being selected for showing an interest in their environment as this relates to attentiveness and exploratory behaviour it potentially taps in to motivation to search and work. No differences were found between withdrawn dogs and pet dogs in any of the elements which could suggest that withdrawn dogs had not been appropriately selected with the optimal level of the reported traits which appear to be of value, or alternatively it could suggest that the working dogs that are withdrawn develop a more pet dog like temperament. As they fail at work they potentially lose the qualities within "Energy & Interest" and "Responsiveness" because they are no longer succeeding at work or getting rewarded through success, the handlers offer less rewards and motivation drops.

Not all aspects of the PANAS and DIAS were found to be related to withdrawal in this study. This could be due to their irrelevance in terms of performance. Previous research has found that not all aspects of temperament could be linked to success or performance in humans (Seibert & Kraimer, 2001), and in addition to this behaviour tests have been found not to be able to predict all learning outcomes in guide dogs (Goddard, 1986). However, it is perhaps more likely that they interact in a more complicated manner or while they may be important in terms of working dog selection they are less affected by working performance outcome in terms of withdrawal. This suggests a need to further refine constructs of temperament traits and measures of performance, including a breakdown of traits related to success and how they are measured. The existing study only looked at dogs in terms of success (being active in the field), and withdrawn (being taken out of service for behavioural reasons), as well as handler performance scores in training and in the field. While this gives an indication of factors related to withdrawal and potential performance levels, it does not look into the dogs in detail in terms of their working role and how traits may differ between different working dogs based on their job. The present study also fails to take into account the impact of breed and sex, as the study populations used were largely mixed. Factors such as sex and breed have been shown to have an impact on service dogs as highlighted in Wilsson and Sundgren, 1996, and differences have been identified in the temperament of dogs in terms of impulsivity using the DIAS depending on their breeding for work or show even within the same breed (Fadel *et al*, 2016). This highlights the complexity of individual differences, and how even looking at a breed level may be insufficient for teasing out behavioural tendencies. While important to consider these factors, the importance of the dog as an individual underpins all of this.

While questionnaire measures are useful they are not always suitable for use, especially in working dogs where often there is not a suitable care giver to complete the questionnaire for the dog, this created a need to develop behavioural tests to measure the temperament components of impulsivity and core affect found to be related to working dog withdrawal. The spatial discounting task developed in Chapter 4 would not be suitable for assessing the reported traits as it could not provide enough detail on the elements identified as correlated with success or withdrawal. The analysis of the working dog population revealed the need for a wider assessment of the behaviour that emerges from the combination of these reported traits and the elements within them. This created the need to develop more refined tests.

7.6 Development of behaviour tests that can distinguish between dogs matching a successful and withdrawn working dog profile.

Behaviour tests are commonly used in working dog assessment as they attempt to emulate working situations and analyse behaviour looking for favourable responses that would be of benefit in an operational context (Asher *et al*, 2013). Questionnaires are generally not used because they cannot be relied upon to be completed effectively and without bias. Knowing the “Responsiveness” and “Energy & Interest” components are of relevance in terms of withdrawal in working dogs, to make this information useful in terms of selection, behavioural assessments able to measure these components are necessary. This led to the development of a series of behavioural assessments to attempt to identify measures that could be used to predict the relevant PANAS and DIAS scores by looking at the elements that are used to generate the components of interest.

7.6.1 Practical Tests

The tests examined learning, response to novelty and sensitivity to social stimuli and were found to have varying degrees of use in terms of differentiating between dogs with similar 'Responsiveness' and "Energy & Interest" profiles to either the working police dogs in active service from the previous chapter, or the withdrawn dogs. The potentially useful tests from the battery include measuring the rate of learning in the discrimination test. While other tests showed interesting results they were not directly comparable between the groups and differences at re test were observed on a number of occasions for measures within the tests.

Where a test is repeated and the response is similar in the two tests, such as the trials taken for the higher scoring group to learn the discrimination task, it indicates the test is robust. However, when the result changes (such as in the discrimination test in terms of the rates of learning which only changed for the lower scoring group between test and retest) there may be a temptation to reject the test as being unreliable, but there is another consideration, which is that the change may reflect learning changes and this change itself may be a measure of interest. It may be that while differences are not present in task performance between the groups, it is the learning process itself that differs and this could be used as a measure in assessment. It could also indicate that differences in training in terms of speed of acquisition may not necessarily predict differences in performance once the task is learned. This is an insight highlighted by this research which could be of importance when assessing behaviour. It could indicate that while rates of learning could act as an indicator of temperament and individual differences, assessing performance in learned and known behaviours may not be able to differentiate in the same way. This is an important consideration in working dogs as during certification the dogs are assessed on known behaviours and how they perform at the tasks they have learned as a measure of ability. But by adding in the training of a new task such as that designed in this project, in such assessments it could give an indication of the underlying temperament components of interest which have been linked to withdrawal in the present study.

In terms of the results of an individual or group performing differently at test compared to retest, animals don't always exhibit the same behaviours when presented with the same situation or context, while this may be a result of habituation or learning, it is important to consider the idea that some animals may be more predictable than others when repeatedly faced with the same situations, which has previously been observed in crustaceans and aquatic species (Stamps *et al*, 2012).

It appears that behavioural differences can be observed in the discrimination test between dogs with different levels of the trait impulsivity in terms of "Responsiveness" and core affect in terms of "Energy & Interest", but the test does not have re-test reliability potentially due to the effects of learning mechanisms as discussed. A great deal more investigation into the robustness and development of the tests and an application to working dogs at selection and in service would be needed to make this work useful as an additional aid to working dog selection.

7.7 General Discussion

This research expands on previous work looking at impulsivity in pet dogs (Wright *et al*, 2011, Wright *et al*, 2012, Reimer *et al*, 2016). "Responsiveness" and "Energy & Interest" as described within this project may be picking up on elements already measured in working dog tests for example the desire for work, which has been shown as being important in predicting certification in detection dogs in Japan (Maejima *et al*, 2007). Desire for work as assessed within Maejima *et al*, 2007 may overlap with responsiveness elements 'my dog is easy to train', 'my dog reacts very quickly', and 'my dog is very interested in new things and new places', it may also overlap with "Energy & Interest" elements 'your dog is full of energy' and 'your dog takes a long time to lose interest in new things'. A combination of these elements reflecting trainability, an interest in the environment and quick responses could result in the appearance of a dog which is eager to work. While it is likely a necessity that working dogs possess a desire to work and this may carry them through selection, to succeed in the field it is also important that the dog remains able to work despite stress.

Caution should be taken when interpreting correlations between temperament and working status, as while it may be the case that certain levels of reported temperament traits may result in withdrawal, it is also possible that the link is effect rather than cause, as temperamental elements may in fact alter when a dog starts to perform poorly in the field. For example, within the "Responsiveness" elements such as 'your dog is full of energy' may go down if the dog is struggling in the field potentially because they are not achieving success they lose the motivation or enthusiasm to work. Longitudinal research would be required to track dogs' performance and temperament from selection through to career termination to determine this link.

This research expands on previous knowledge in terms of establishing relationships between temperament components relating to impulsivity and core affect and withdrawal in a population of working police dogs. Furthermore, behavioural assessments were developed to correlate with the questionnaire component scores of relevance in terms of working dog withdrawal. The questionnaires and behavioural tests (with further development and investigation into re test reliability) could go on to be used to establish population norms for working dogs and go on to be used to identify dogs at risk of withdrawal from service and once cause or effect of temperament and withdrawal is understood such measures could be used to inform the selection process. There is a need to follow through from selection to certification and on to end of career to establish the progression of temperament as it links to career lifetime performance. There is also a need to establish the usefulness of such testing in younger dogs for instances where these are being selected for work, as currently the assessment of such temperament components in young dogs is not well established and age has been found to impact behaviour assessments (Slabbert & Odendaal, 1999. Batt et al, 2008). However, it is acknowledged that early life experiences influence the expression of temperament traits, and these traits develop and become more stable with age (Batt *et al*, 2008), this could be the reason why in the spatial discounting task the young dogs performance in the test and questionnaire did not correlate. Behavioural responses of dogs are thought to develop with age as a result of experience and maturation, so the ability to predict adult behaviour in young dogs is thought to be limited. In guide dogs, while temperament has been reported to have usefulness in selecting dogs that will be able to pass certification, the age at which such temperament testing can be reliably implemented is not helpful as the dogs are selected for work much younger than the age at which such traits can be reliably predicted (Asher *et al*, 2013). Little is known about the stability of temperament traits in younger dogs, particularly in the case of using behaviour tests to assess temperament (Knoll *et al*, 1998. Batt *et al*, 2008). The predictive validity of such testing has been shown to improve with age, so only limited conclusions can be drawn at the selection stage (Batt *et al*, 2008). Developing an understanding of the development of temperament with age could enable more specific training programmes to focus on developing the temperament towards a desired optimum on an individual basis, maximising chances of selection and minimising risk of withdrawal.

7.8 Limitations

Within this project there are a number of limitations that should be considered. While the focus of this thesis is on the relationship between core affect and impulsivity, and working dog withdrawal, it is important to acknowledge that performance will be affected by a number of other factors that this project did not take into account. The working dog population within this thesis examined a mixture or patrol and detection dogs and factors that affect these dogs in work could potentially differ (for detection dogs environmental challenges (Haverbeke *et al*, 2008) (including factors such as depth of burial for detection dogs (Diverio *et al*, 2016), climate factors such as temperature extremes (Tiffoli & Roilfe, 2006) and wind direction (Diverio *et al*, 2016) may have a greater impact), although some factors may be similar across disciplines such as physical fitness levels (Gazit and Terkel, 2003), age (Tiffoli & Rolfe (2006), and training contexts (Gazit *et al*, 2005). Other factors include, but are not limited to, handler error or interference (Curran *et al*, 2010), the personality of the handler (Kotrschal *et al*, 2009) and handler stress (Zubedat *et al*, 2014). This highlights the issue of not breaking down analysis into working dogs role and field of regular deployment, which was prevented by small sample sizes. This leaves information unknown as to if or how traits differ between dogs of different disciplines. Furthermore, while information was obtained about active and withdrawn police dogs, only information about active military working dogs was obtained in this study, limiting the conclusions that can be drawn about withdrawn military working dogs. While the active military dogs were statistically similar to the active police dogs, there is a need to profile withdrawn military working dogs as they may differ in their psychometric profiles to withdrawn police dogs, in which case different tests would need to be developed to assess these differences. A number of other factors can affect working dog performance such as dog handler relationship has also been shown to impact on performance (Lefebvre *et al*, 2007), and dogs with single handlers performing better than those worked by multiple handlers (Nolan & Gravitte, 1977). Dogs have also been shown to be impacted by unintentional handler cues directed at targets (Wasser *et al*, 2004) as it has been shown that some working dogs will attend to human cues over vision and olfaction so could be compromised by handler error (Szetei *et al*, 2003). This highlights that no matter how robust the selection process, aspects of the environment, training and handling will always confound the final assessment of the dog in work.

Within this project, pet dogs were used in a number of the studies, rather than working dog populations due to sample availability. While this could be seen to pose an issue in terms of the viability of the dogs used in the sample, the dogs were selected for specific temperament profiles that matched those of the working dog study to minimise this effect in Chapter 6. As well as this the temperament traits measured using the PANAS and the DIAS have been shown to be relatively stable throughout life (Reimer *et al*, 2016), therefore experiences and training are unlikely to change the way they appear when adult dogs are assessed using these questionnaires. By tapping into this level of temperament, it is arguable that it does not matter what role the dog has, pet or working. Variation within the temperament of the pet dog population allowed for the selection of pet dogs for the study in Chapter 6 with temperament profiles matching those of the working dogs in Chapter 5.

Relatively small sample sizes were used throughout this project, which does limit the conclusions that could be drawn and the level of analysis that could be undertaken. While this limits conclusions in terms of working role, the study still provides useful information in terms of factors correlated to withdrawal in general in working dogs.

As discussed in the introduction, questionnaires can provide valuable information on dog behaviour, especially as a quick form of assessment where behavioural testing may not be practical or possible. They do have their limitations though, as they rely on a consistent person interacting with the dog providing accurate information on that dog's behaviour in a range of contexts. Moreover, validation is also an issue and in this research the PANAS and the DIAS used have not been validated for use in working dogs. This could mean there are issues with the interpretation of the questions in a working dog sector which could impact the results as dog handlers may have a different understanding or interpretation of their dog's behaviour in relation to the questions asked. However, the questionnaires have been validated in large populations of pet dogs. While differences between populations of dogs such as pet dogs, show dogs and working dogs have been identified previously (Weiner *et al*, 2016. Serpell & Hsu, 2005), the interpretation of behaviour has only been identified as an issue between cultures, not different levels of dog owning or handling experience (Wan *et al*, 2009. Serpell & Hsu, 2005).

Within the working dog study, collecting information on PANAS and DIAS for withdrawn dogs relied on handler recall of the dog from up to one year previously. While the 1 year time limit was requested to limit the issues of memory recall, there is still the potential for recall bias in these results. Campbell *et al,* 2011 found that in humans new information was provided in a memory recall test one year on, which could suggest that memory improves over time, or that events are not being remembered correctly, either way it could indicate a lack of reliability in memory. Recall bias could only be avoided by obtaining results at the moment a dog is withdrawn, which ties back into the need to track dogs longitudinally from pre-selection through to withdrawal or end of service.

All of the studies within this thesis did not preselect dogs based on age, sex or breed, sample had to be opportunistic for the most part to enable swift participation within the timeframe of the project. Where this was not the case, dogs were being selected based on specific temperament profiles form the PANAS and DIAS, which again had to be recruited once identified rather than waiting for all of the dogs to match the same age, sex and breed due to time constraints. While it is acknowledged that such mixed samples could have impacted on the results (as discussed earlier, such as sex and breed impacting service dogs as highlighted in Wilsson and Sundgren, 1996), this opportunistic sampling was necessary to complete the project within the allocated timeframe. While it is recognised that factors such as breed may impact on temperament, the focus within this project is on each dog as an individual beyond factors such as breed. However, it is important to acknowledge that age is still a factor that needs to be taken into account with assessments mainly in respect to young dogs. The impulsivity work in Chapter 4 was unable to provide reliable behavioural measures for dogs aged 2-9 months so further work is needed to establish at what point the behavioural test correlates with questionnaire score, or to develop another non-questionnaire method of assessment to enable the earliest possible assessment.

7.9 Future Work

There is a need to longitudinally assess young dogs through to adulthood in terms of the behaviour test for impulsivity to identify how the trait develops and at which point it becomes stable enough to behaviourally measure the finer elements within the trait such as "Responsiveness". While core affect questionnaire measures have been validated in dogs from 10 weeks of age behaviourally this has not been assessed so behavioural examination of the elements of relevance would need to be studied in a population of young dogs and tracked to identify when the measures become stable or any trends that appear. This will not only provide clarity in terms of the trait itself but also facilitate the use of the behavioural test to accurately measure the trait from as young an age as possible, furthering the practical applications particularly in terms of identifying dogs at risk of having behavioural problems (Taylor & Mills, 2006).

Longitudinal work is also required on working dog temperament to understand cause and effect in relation to temperament and success or withdrawal. By using the psychometric profiling system from selection through to contract termination, it could be possible to identify whether it is the dog's temperament itself that results in eventual withdrawal, or whether the dog's temperament alters as a result of the dog failing at work. If the latter there may be possible interventions to prevent the dog from being withdrawn from service, so this could act as an early warning system and prevent mistakes being made in the field. It may also be possible to mitigate deficiencies in temperament that exist before training using remedial training to focus on the deficient areas.

Temperament traits related to core affect and impulsivity were found to be related to withdrawal from service in a population of UK police dogs. Significant factors relating to withdrawal of dogs from service can be used to consider individuals currently working that are at risk of withdrawal or not performing at their best. Further work is needed to determine cause or effect in relation to the traits found to be related to withdrawal from service and there is a need to further develop and refine behaviour tests to measure these traits to show convergent validity and ensure the traits can be measured in a practical way. During the selection process the discrimination test and questionnaires used within this project could be conducted alongside the selection and certification processes, as well as at withdrawal to establish if or how performance changes over the lifetime of the working dog to establish if temperament has a causal or effective role in working dog success or withdrawal. By running the profiling and test developed in this thesis alongside current practices it could be identified how the tests could be used to predict successful or withdrawn outcomes. Furthermore a larger sample of working dogs both in active service and especially those withdrawn from service could be used to develop a scale as to where the risk of withdrawal from scores on the questionnaires makes it better to not purchase or select the dog.

7.10 Recommendations

Recommendations from this work to the working dog sector need to be cautious. The PANAS and DIAS could be used as additional screening tools on borderline decision dogs who either almost or only just meet working dog selection requirements to help inform decisions. Adding these tools into the selection processes and then following up dogs could help to validate them as an assessment procedure for identification of dogs unsuitable for work.

7.11 Final Conclusions

This thesis has highlighted important issues relating to the selection of working dogs, ranging from inconsistencies in terminology to potential gaps within the selection process itself. It highlights the need to examine traits outside of fear and confidence in working dog selection and has highlighted the potential working dog applications of temperament profiling tests used in pet dogs to identify individuals at risk of behavioural problems. Relationships between elements of temperament and working status have been identified, although cause and effect is still to be determined, questionnaires can be used to identify differences between active and withdrawn working dog profiles. While arousal and performance may be linked on an individual basis, further work developing methodologies would be necessary to identify the importance of this for the field as well as how this knowledge can be used to optimise working dog usefulness. This project has increased our understanding of working dogs, but it has also raised further questions, particularly relating to the nature of the relationship (cause or effect) between temperament elements from the PANAS and DIAS and working dog status (active or withdrawn). It is important to strive to understand the temperament traits which predict lifelong success as well as to develop methods to predict it from an early age.

Acknowledgements

Project funded by Dstl

Supervisory Team:

Daniel Mills

Helen Zulch

Hannah Wright

External Supervisor:

Nina Cracknell

I would personally like to thank Daniel Mills & Helen Zulch for their continued support and guidance during this project.

Special thanks to Lynn Hewison for her work on the Spatial Discounting Task, as well as Naomi Mitchell, Laura Adkins, Martje Smesters, Lisa Dietz, Kate Ellam and Sophie Moszuti for assisting with data collection during various stages of the project. Extra special thanks to Pam MacKinnon of Talking Dogs and all of the super sniffers who participated in the scent detection task.

The author would also like to thank all of the dogs, owners and handlers for taking part in the research, and Greater Manchester Police for their role in data collection.

Ethical Approval

Ethical Approval code: UID CoSREC124

References

Ainslie, G. (1974). Impulse control in pigeons. *Journal of Experimental Animal Behaviour*. 21, p485-489

Anderson, K. (1994). Impulsivity, caffeine, and task difficulty: A within-subjects test of the Yerkes-Dodson law. *Personality and Individual Differences.* 16(6), 813-829.

Anderson, K. J., Revelle, W., & Lynch, M. J. (1989). Caffeine, impulsivity, and memory scanning: A comparison of two explanations for the Yerkes-Dodson Effect. *Motivation and Emotion*, 13, 1-20.

Arata, S. Momozawa, Y. Takeuchi, Y. Mori, Y. (2010). Temperament assessment on guide dogs using questionnaire survey and behaviour tests. *Journal of Veterinary Behaviour*. 5 (1), 58.

Arnott, E. R., Early, J. B., Wade, C. M., McGreevy, P. D. (2014a). Estimating the economic value of Australian stock herding dogs. *Animal Welfare, 23*, 189–197.

Arnott, E. R., Early, J. B., Wade, C. M., McGreevy, P. D. (2014b). Environmental factors associated with success rates of Australian stock herding dogs. *PLOS ONE 9*, e104457.

Asher, L. Blythe, S. Roberts, R. Toothill, L. Craigon, P. Evans, K. Green, M. Englang, G. (2013). A standardised behaviour test for potential guide dog puppies: Methods and association with subsequent success in guide dog training. *Journal of Veterinary Behaviour*. 8 (1), p431-438

Barnard, S. Siracusab, C. Reisner, I. Valsecchi, P, Serpell, J. (2012). Validity of model devices used to assess canine temperament in behavioral tests. *Applied Animal Behaviour Science*. 138 (1-2), 79–87.

Barrick, M. Parks, L. Mount, M. (2005). Self-monitoring as a moderator of the relationships between personality traits and performance. *Personnel Psychology*. 58 (1), 745-767.

Batt, L. Batt, M. Baguley, J. McGreevy, P. (2008). Factors associated with success in guide dog training. *Journal of Veterinary Behaviour*. 3 (1), p143-151.

Bell, A. (2007). Future directions in behavioural syndromes research. *Biological Sciences*. Proc. R. Soc. B 274, 755–761

Bickel, W. Miller, M. Yi, R. Kowal, B. Lindquist, D. Pitcock, J. (2007). Behavioural and neuroeconomics of drug addiction: Competing neural systems and temporal discounting processes. *Drug and Alcohol Dependence.* 90(1), p85-91.

Bindra, D. Spinner, N. 1958Response to Different Degrees of Novelty: The Incidence of Various Activities. Journal of the Experimental Analysis of Behavior. 1(4), 341-350. DOI: 10.1901/jeab.1958.1-341

Bollen, K. Horowitz, J. (2008). Behavioral evaluation and demographic information in the assessment of aggressiveness in shelter dogs. *Applied Animal Behaviour Science*. 112 (1-2), 120-135.

Bower, G. 1981. Mood and memory. *American Psychology*. 36 (1), 129-148

Bray, E. MacLean, E. Hare, B. (2015). Increasing arousal enhances inhibitory control in calm but not excitable dogs. *Animal Cognition*. 18(6):1317-29. doi: 10.1007/s10071-015-0901-1

Burke, M. Mathews, A. 1992. Autobiographical memory and clinical anxiety. *Cognitive Emotion*, 6 (1), 23-35

Burman, O. McGowan, R. Mendl, M. Norling, Y. Paul, E. Rehn, T. Keeling, L. 2011. Using judgement bias to measure positive affective state in dogs. *Applied Animal Behaviour Science.* 132(3-4), 160-168.

Burns, K. (2016). Inside the headquarters for military dogs. *JAVMA, 248*(11), 1199-1203.

Cohen, J. (1960). A coefficient of agreement for nominal scales*. Educational and Psychological Measurement, 20,* 37- 46.

Call, J. Bräuer, J. Kaminski, J. Tomasello, M. 2003. Domestic dogs (*Canis familiaris*) are sensitive to the attentional state of humans. *Journal of Comparative Psychology*, 117(3), 257-263.

Call, J. Carpenter, M. 2001. Do apes and children know what they have seen? *Animal Cognition*, 4(1), 201-220.

Campbell, J. Nadel, L. Duke, D. Ryan, L. 2011. Remembering all that and then some: Recollection of autobiographical memories after a 1-year delay. *Memory*. 19(4), 406-415

Capitainio, J. Blozis, S. Snarr, J. Steward, A. McCowan, B. (2005). Do Birds of a Feather Flock Together or Do "Opposites Attract"? Behavioural Responses and Temperament Predict Success in Pairings of Rhesus Monkeys in a Laboratory Setting. *American Journal of Primatology*. 1 (1), 1-11.

Carrier, L. Cyr, A. Anderson, R, Walsh, C, 2013. Exploring the dog park: relationships between social behaviours, personality and cortisol in companion dogs. Applied Animal Behaviour Science. 146(1), 96-106.

Chamorro-Premuzic, T. Furnham, A. (2003). Personality Traits and Academic Examination Performance. *European Journal of Personality*. 17 (1), 237-250.

Clark, R. Squire, L. 1998. Classical conditioning and brain systems: the role of awareness. *Science*, 280 (1), 77-81.

Cohen, R. (2011). Yerkes–Dodson Law. Encyclopedia of Clinical Neuropsychology. DOI 10.1007/978-0-387-79948-3_1340. Print ISBN 978-0-387-79947-6. Online ISBN 978-0-387-79948-3. Springer New York. Springer Science+Business Media, LLC. pp 2737-2738

Coleman, K. Tully, L. McMillan, J. (2005). Temperament Correlates with Training Success in Adult Rhesus Macaques. *American Journal of Primatology*. 65 (1), 63-71.

Colom, R. Escorial, S. Chun Shih, P. Privado, J. (2007). Fluid intelligence, memory span, and temperament difficulties predict academic performance of young adolescents. *Personality and Individual Differences*. 42 (1), 1503-1514.

Critchfield, T. Kollins, S. (2001). Temporal discounting: Basic research and the analysis of socially important behaviour. *Journal of Applied Behaviour Analysis.* 34(1), p101-122

Curran, A. Prada, P. Furton, K. (2010). Canine human scent identifications with post-blast debris collected from improvised explosive devices. *Forensic Science International.* 199(1), 103-108.

Cunningham, W. Raye, C. Johnson, M. 2004. Implicit and explicit evaluation: fMRI correlates of valence, emotional intensity, and control in the processing of attitudes. *Journal of Cognitive Neuroscience*. 16(10), 1717– 1729. [PubMed: 15701224].

Danner, U. Aarts, H. de Vries, N. (2008). Habit vs. intention in the prediction of future behaviour: The role of frequency, context stability and mental accessibility of past behaviour. *British Journal of Social Psychology.* 47(1), 245-265.

Dawkins, M. (2004). Using behaviour to assess animal welfare. *Animal Welfare*. 13 (1), 3-7.

Deidrich, C. Giffroy, J. 2006. Behavioural testing in dogs: A review of methodology in search of standardisation. *Applied Animal Behaviour Science.* 97(1), 51-72.

De Meester, R. De Bacquwr, D. Peremans, K. Vermeire, S. Planta, D. Coopman, F. Audenaert, K. (2008). A preliminary study on the use of the Socially Acceptable Behaviour test as a test for shyness/confidence in the temperament of dogs. *Journal of Veterinary Behaviour*. 3 (1), p161-170.

De Meester, R. Pluijmakers, J. Vermeire, S. Laevens, H. (2011). The use of the Socially Acceptable Behaivour test in the study of temperament of dogs. *Journal of Veterinary Behaviour*. 6 (1), p211-224.

De Palma, C. Vigganio, E. Barillari, W. Palme, R. Dufour, A. Fantini, C. Natoli, E. (2005). Evaluating the temperament in shelter dogs. *Behaviour*. 142 (1), 1307-1328.

Derryberry, D. Reed, M. (1994). Temperament and Attention: Orienting Toward and Away from Positive and Negative Signals. *Journal of Personality and Social Psychology*. 66 (6), 1128-1139.

Desiré, L. Boissy, A. Veissier, I. 2002. Emotions in farm animals: a new approach to animal welfare in applied ethology. *Behavioral Processes*. 60 (1), 165-180

Dimitriou, L. Sharp, N. Doherty, M. (2002), Circadian effects on the acute responses of salivary cortisol and IgA in well trained swimmers. *British Journal of Sports Medicine*. 36 (1), p260-264.

Dingemase, N. Reale, D. (2005). Natural selection and animal personality. *Behaviour*. 142 (2005), 1159-1184

Dittmar, H. Bond, R. (2010). 'I want it and I want it now': Using a temporal discounting paradigm to examine predictors of consumer impulsivity. *British Journal of Psychology.*101 (1), p751-776

Diverio, S. Barbato, O. Cavallina, R. Guelfi, G. Iaboni, M. Zasso, R. Di Mari, W. Santoro, M. Knowles, T. (2016). A simulated avalanche search and rescue mission includes temporary physiological and behavioural changes in military dogs. *Physiology and Behaviour.* 163(1), 193-202.

Doelling, J. Johnson, J. (1990). Predicting success in foster placement: The contribution of parent-child temperament characteristics. *American Journal of Orthopsychiatry*. 60 (4), 585-593.

Dreschel, N. Granger, D. (2009). Methods of collection for salivary cortisol measurements in dogs. *Hormones and Behaviour*. 55 (1), p163-168.

Duffy, D. Kruger, K. Serpell, J. 2014. Evaluation of a behavioral assessment tool for dogs relinquished to shelters. Preventative Veterinary Medicine. 117(1), 601-609.

Duffy, D. Serpell, J. (2012). Predictive validity of a method for evaluating temperament in young guide and service dogs. *Applied Animal Behaviour Science*. 138 (1), 99-109.

Evans, T. Beran, M. (2007). Chimpanzees use self-distraction to cope with impulsivity. *Biology Letters*. 3, p599-602

Evans, R. Herbold, J. Bradshaw, B. Moore, G. (2007). Causes for discharge of military working dogs from service: 268 cases (2000-2004). *Journal of the American Veterinary Medical Association.* 231(8), 1215-1220.

Evenden, J. L. (1999). "Varieties of impulsivity". *Psychopharmacology.* 146 (4): 348–61.

Eysenck, M. Mogg, K. May, J. Richards, A. Mathews, A. 1991. Bias in interpretation of ambiguous sentences related to threat and anxiety. *Journal of Abnormal Psychology*. 100 (1), 144-150

Fadel, F. Mills, D. Pilot, M. (2016). Assessing traits for behavioral genetics studies. *Canine Behaviour and Genetics.* Accessed January 2016.
https://www.researchgate.net/profile/Fernanda_Fadel/publication/281578383_Assessing_traits_for_behavioural_genetics_studies/links/5742df6708ae9f741b379091.pdf

Feldman Barrett, L., & Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. Journal of Personality and Social Psychology, 74(4), 967-984.

Fidler, A. E., van Oers, K., Drent, P. J., Kuhn, S., Mueller, J. C., & Kempenaers, B. (2007). Drd4 gene polymorphisms are associated with personality variation in a passerine bird. *Proceedings of the Royal Society B*, **274**, 1685–1691.

Folkman, S. Lazarus, R. Dunkel-Schetter, C. DeLongis, A. Gruen, R. (1986). Dynamics of a stressful encounter: Cognitive appraisal, coping, and encounter outcomes. *Journal of Personality and Social Psychology*, 50(5), 992-1003. http://dx.doi.org/10.1037/0022-3514.50.5.992

Fox, M. (1971). *Behaviour of wolves dogs and related canids*. Dogwise Classics. ISBN: 9781617810480.

Foyer, P. Bjällerhag, N. Wilsson, E. Jensen, P. (2014). Behaviour and experiences of dogs during the year of life predict the outcome in a later temperament test. *Applied Animal Behaviour Science*. 155 (1), 93-100.

Foyer, P, Svedberg, A. Nilsson, E. Wilsson, E. (2015). Behaviour and cortisol responses of dogs evaluated in a standardised temperament test for military working dogs. *Journal of Veterinary Behaviour.* Article in Press, 1-6.

Foyer, P. Svedberg, A. Nilsson, E. Wilsson, E. Faresjo, A. Jensen, P. (2016). Behaviour and cortisol responses of dogs evaluated in a standardised temperament test for military working dogs. *Journal of Veterinary Behavior.* 11(1), 7-12.

Foyer, P. Wilsson, E. Wright, D. Jensen, P. (2013). Early experiences modulate stress coping in a population of German shepherd dogs. *Applied Animal Behaviour Science.* 146(1), 79-87.

French, J. (1993). Assessment of donkey temperament and the influence of home environment. *Applied Animal Behaviour Science.* 36 (1), 249-257.

Fuchs, T. Gaillard. C, Gebhardt-Henrich, S. Ruefenacht, S. Steiger, A. (2005). External factors and reproducibility of the behaviour test in German shepherd dogs in Switzerland. *Applied Animal Behaviour Science*. 94 (1), p287-301.

Furton, K. Myers, L. (2001). The scientific foundation and efficacy of the use of canines as chemical detectors for explosives. *Talanta.* 54(1), 487-500.

Gabora, L., & Colgan, P. (1990). A model of the mechanisms underlying exploratory behavior. In (S. Wilson & J. A. Mayer, Eds.) *Proceedings of the First International Conference on the Simulation of Adaptive Behavior,* 475-484, Cambridge MA: MIT Press.

Garner, J. Meehan, C. Mench, J. 2003. Stereotypies in caged parrots, schizophrenia and autism: evidence for a common mechanism. *Behavioral Brain Research*. 145 (1), 125-134

Gartner, M. (2015). Pet personality: A review. P*ersonality and Individual Differences*. 75 (1), 102–113.

Gerber, A. Posner, J. Gorman, D. Colibazzi, ,C. Yu, S. Wang, Z. Kangarlu, A. Zhu, H. Russell, J. Peterson, B. 2008. An affective circumplex model of neural systems subserving valence, arousal, and cognitive overlay during the appraisal of emotional faces. *Neuropsychologia.* 46(8), pp2129-2139

Goddard, M. Beilharz, R. (1986). Early prediction of adult behaviour in potential guide dogs. *Applied Animal Behaviour Science*. 15 (1), 247-260.

Goldberg, L. (1993). The Structure of Phenotypic Personality Traits. *American Psychologist*. 48 (1), 26-34.

Gorecka-Bruzda, A. Jezierski, T. (2010). Breed differences in behaviour-related characteristics of stallions evaluated in performance tests. *Animal Science Papers and Reports.* 28(1), 27-36.

Gosling, S. (2008). Personality in non-human animals. *Social and Personality Psychology Compass* 2/2 (2008): 985–1001, 10.1111/j.1751-9004.2008.00087.x

Gosling, S. (2001). From mice to men: What can we learn about personality from animal research? *Psychological Bulletin*. 127 (1), 45-86.

Gosling, S. John, O. (1999). Personality Dimensions in Nonhuman Animals: A Cross-Species Review. *Current Directions in Psychological Science*. 8 (3), 69-75.

Gosling, S. Vazire, S. (2002). Are we barking up the right tree? Evaluating a comparative approach to personality. *Journal of research in Personality*. 36 (1), 607-614.

Grazit, I. Goldblatt, A. Terkel, J. 2005. The role of context specificity in learning: The effects of training context on explosives detection in dogs. *Animal Cognition.* 8(1), 143-150.

Grazit, I. Terkel, J. (2003). Explosives detection by sniffer dogs following strenuous physical activity. *Applied Animal Behaviour Science.* 81(1), 149-161.

Green, L. Myerson, J. Holt, D. Slevin, J. Estle, S. (2004). Discounting of delayed food rewards in pigeons and rats: Is there a magnitude effect? *Journal of the Experimental Analysis of Behavior*. 81 (1), 39-50.

Hagvet, K. Hanin, Y. (2007). Consistency of performance-related emotions in elite athletes: Generalizability theory applied to the IZOF model. *Psychology of Sport and Exercise*. 8 (1), p47-72.

Hall, S. S., MacMichael, J., Turner, A., & Mills, D. S. (2017). A survey of the impact of owning a service dog on quality of life for individuals with physical and hearing disability: a pilot study. *Health and Quality of Life Outcomes*, *15*(1), 59.

Hall, S., Dolling, L., Bristow, K., Fuller, T., & Mills, D. S. (2016). *Companion animal economics: the economic impact of companion animals in the UK*. CABI.

Harvey, N. Craigon, P. Sommerville, R. McMillan, C. Green, M. England, G. Asher, L. 2016. Test-retest reliability and predictive validity of a juvenile guide dog behaviour test. *Journal of Veterinary Behaviour: Clinical Applied Research*. 11(1), 65-76.

Haverbeke, A. Deiderich, C. Depiereux, E. Giffroy, J. (2008). Cortisol and behavioural responses of working dogs to environmental challenges. *Physiology and Behaviour.* 93(1), 59-67.

Haverbeke, A. De Smet, A. Depiereux, E. Giffroy, J. Deiderich, C. (2009). Assessing undesired aggression in military working dogs. *Applied Animal Behaviour Science*. 117 (1-2), 55-62.

Herry, C. Bach, C. Esposito, F. Di Salle, F. Perrig, W. Scheffler, K. Luthi, A. Seifritz, E. 2007. Processing of temporal unpredictability in human and animal amygdala. *Journal of Neuroscience*. 27 (1), 5958-5966

Horwitz, D. Mills, D. 2009. BSAVA Manual of Canine and Feline Behavioural Medicine. Second Edition. British Small Animal Veterinary Association. Gloucester. ISBN 978 1 905319 15 2. 270-272.

Horvath, Z. Igyarto, B. Magyar, A. Miklosi, A. 2007. Three different coping styles in police dogs exposed to a short term challenge. *Hormones and Behaviour.* 52(1), 621-630.

Hsu, Y. Serpell, J. 2003. Development and validation of a questionnaire for measuring behaviour and temperament traits in pet dogs. *Journal of American Veterinary Medical Association*. 223(1). 1293-1300.

Hsu, Y. Sun, L. 2010. Factors associated with aggressive responses in pet dogs. *Applied Animal Behaviour Science*. 123(1), 108-123.

Huang, C. Hung, T. (2010). Using the IZOF hypothesis to examine the relationship between salivary cortisol and shooting performance. *Abstracts/International Journal of Physchology*. 77 (1), p274.

Huntingford, F. (1976). The relationship between anti-predator behaviour and aggression among conspecifics in the three spined stickleback, *Gasterosteus aculeatus. Animal Behaviour.* 24 (1), 245-260.

Ilska, J. Haskell, M. Blott, S. Sánchez-Molano, E. Polgar, Z. Lofgren, S. Clements, D. Wiener, P. 2017. Genetic characterisation of dog personality traits. *Genetics*. Early Online, published on April 10, 2017 as 10.1534/genetics.116.192674.

Jonckheer-Sheehy, V. Vinke, C. Ortolani, A. (2012). Validation of a Polar human heart rate monitor for measuring heart rate and heart rate variability in adult dogs under stationary conditions. *Journal of Veterinary Behaviour*. 7 (1), p205-212.

Jones, A. Gosling, S. (2005). Temperament and personality in dogs (Canis familiaris): A review and evaluation of past research. *Applied Animal Behaviour Science*. 95 (1-2), 1–53.

Judge, T. Ilies, R. (2002). Relationship of Personality to Performance Motivation: A Meta-Analytic Review. *Journal of Applied Psychology*. 87 (4), 797-807.

Kato, M. Miyaji, K. Ohtani, N. Ohta, M. 2012. Effects of prescription diet on dealing with stressful situations and performance of anxiety-related behaviors in privately owned anxious dogs. *Journal of Veterinary Behaviour: Clinical Applied Research* 7(1), 21-26.

Kobayashi, N. Arata, S. Hattori, A. Kohara, Y. Kiyokawa, Y. Takeuchi, Y. Mori, Y. 2013. Association of puppies' behavioural reaction at five months of age assessed by questionnaire with their later 'Distraction' at 15 months of age, an important behavioural trait for guide dog qualification. *Journal of Veterinary Medical Sciencve.* 75(1), 63-67.

Konok, V. Kosztolanyi, A. Rainer, W. Mutschler, B. Halsband, U. Miklosi, A. 2015. Influence of owners' attachment style and personality on their dogs' (Canis familiaris) separation-related disorder. *PLoS One*. 10(1), 0118375.

Knol, N. 2001Roozendaal, C. Vandenbogaard, L. Bouw, J. (1988). The suitability of dogs as guide dogs for the blind: Criteria and testing procedures. *Veterinary Quarterly.* 10 (10), 198-204.

Komar, D. (1999). The use of cadaver dogs in locating scattered, scavenged human remains: preliminary field test results. *Journal of Forensic Science.* 44(1), 405-408.

Komarraju, M. Karau, S. Schmeck, R. (2009). Role of the Big Five personality traits in predicting college students' academic motivation and achievement. *Learning and Individual Differences*. 19 (1), 47-52.

Koolhaas, J. Korte, S. De Boer, S. Van Der Vegt, B. Van Reenen, C. Hopster, H. De Jong, I. Ruis, M. Blokhuis, H. 1999. Coping styles in animals: current in behaviour and stress-physiology. *Neuroscience and Behavioural Reviews.* 23(1), 925-935.

Kotrschal, K. Schoberl, I. Bauer, B. Thibeaut, A. Wedl, M. (2009). Dyadic relationships and operational performance of male and female owners and their male dogs. *Behavioural Processes.* 81(1), 383-391.

Kruk, M. (1997). Measuring behaviour into the twenty-first century. Measuring Behaivour Meeting Report 1996. *Trends in Neuroscience.* 20(5), 187-189.

Kubinyi, E. Topal, J. Miklosi, A. Csanyi, V. 2003. Dogs (*Canis familiaris*) learn their owners via observation in a manipulation task. *Journal of Comparative Psychology.* 117(2), 156-165

Kubinyi, E. Turcsan, B. Miklosi, A. 2009. Dog and owner demographic characteristics and dog personality trait associations. *Behavioural Processes*. 81(1), 392-401.

Lane, R. McRae, K. Reiman, E. Chen, K. Ahern, G. Thayer, J. (2009). Neural correlates of heart rate variability during emotion. *NeuroImage*. 44 (1), p213-222.

Lane, R. Chua, P. Dolan, R. 1999. Common effects of emotional valence, arousal and attention on neural activation during visual processing of pictures. *Neuropsychologia*. 37(9), 989-997

Lautenbach, F. Laborde, S. Achtzehm, S. Raab, M. (2014). Preliminary evidence of salivary cortisol predicting performance in a controlled setting. *Psychoneuroendocrinology*. 42 (1), p218-224.

Lazarowski, L., & Dorman, D. C. (2014). Explosives detection by military working dogs: olfactory generalization from components to mixtures. *Applied Animal Behaviour Science*, *151*, 84-93.

Ledger, R. A., & Baxter, M. R. (1997). *The development of a validated test to assess the temperament of dogs in a rescue shelter*. In: Proceedings of the First International Conference on Veterinary Behavioural Medicine, Birmingham, UK, 87–92.

Lefebvre, D. Diederich, C. Delcourt, M. Giffroy, J. (2007). The quality if the relation between handler and military dogs influences efficiency and welfare of dogs. *Applied Animal Behaviour Science.* 104(1), 49-60.

Lefebvre, D. Giffroy, J. Diederich, C. 2009. Cortisol and behavioural responses to enrichment in military working dogs. *Journal of Ethology.* 27(1), 255-265.

Lerner, J. Keltner, D. 2000. Beyond valence: toward a model of emotion-specific influences on judgement and choice. *Cognitive Emotion*, 14 (1), 473-493

Lewis, P. Critchley, H. Rotshtein, P. Dolan. 2007. Neural Correlates of Processing Valence and Arousal in Affective Words. Cerebral Cortex. 17(3). Pp742-748

Ley, J., Bennett, P., & Coleman, G. (2008). Personality dimensions that emerge in companion canines. *Applied Animal Behaviour Science*, *110,* 305–317.

Ley, J. Coleman, G. Holmes, R. Hemsworth, P. 2007. Assessing fear of novel and startling stimuli in domestic dogs. Applied Animal Behaviour Science. 104(1-2), 71-84. https://doi.org/10.1016/j.applanim.2006.03.021.

Lindau, M. Almkvist, O. Mohammed, A. 2016. Effects of Stress on Learning and Memory. Stress: Concepts, Cognition, Emotion, and Behavior, Academic Press, 2016, pp153-160

Long, A. Platt, M. (2005). Decision Making: The Virtue of Patience in Primates. *Current Biology*. 15 (21), 874–876.

Lorenzo, N. Wan, T. Harper, R. Hsu, Y. Chow, M. Rose, S. Furton, K. (2003). Laboratory and field experiments used to identify Canis lupus var. familiaris active odour signature chemicals from drugs, explosives, and humans. *Analytical and Bioanalytical Chemistry.* 376(8), 1212-1224.

Luft, C. Takase, E. Darby, D. (2009). Heart rate variability and cognitive function: Effects of physical effort. *Biological Psychology*. 82 (2), p186-191.

Lyons, D. Price, E. Moberg, G. (1988). Individual differences in temperament of domestic dairy goats: constancy and change. *Animal Behaviour.* 36 (1), 1323-1333.

MacLeod, C. Byrne, A. 1996. Anxiety, depression, and the anticipation of future positive and negative experiences. *Journal of Abnormal Psychology*. 105 (1), 286-289

Maejima, M. Inoue-Murayama, M. Tonosaki, K. Matsuura, N. Kato, S. Saito, Y. Weiss, A. Murayama, Y. Ito, S. (2007). Traits and genotypes may predict the successful training of drug detection dogs. *Applied Animal Behaviour Science*. 107 (1), 287-298.

Maltby, J. Day, L. Macaskill, A. (2013). Personality, Individual Differences and Intelligence. Pearson; 3rd edition, ISBN10: 0131297600

Mann, R. (1959). A review of the relationships between personality and performance in small groups. *Psychological Bulletin*. 56 (4), 241-270.

Marshall-Pescini, S. Valsecchi, P. Petak, I. Accorsi, P. Previde, E. (2008). Does training make you smarter? The effects of training on dogs' performance (*Canis familiaris)* in a problem solving task. *Behavioural Processes.* 78(1), 449-454.

Martin, J. Reale, D. (2008). Temperament, risk assessment and habituation to novelty in eastern chipmunks, *Tamias striatus*. *Animal Behaviour*. 75 (1), 309-318.

Mathews, A. Macleod, C. 1994. Cognitive approaches to emotion and emotional disorders. *Annual Review Psychology.* 45 (1), 25-50

McCrae, R. R., Costa, P. T. Jr, Ostendorf, F., Angleitner, A., Hrebickova, M., Avia, M. D., Sanz, J., *et al*. (2000). Nature over nurture: Temperament, personality, and life span development. *Journal of Personality and Social Psychology*, **78**, 173–186.

McCune, S. (1995). The impact of paternity and early socialisation on the development of cats' behaviour to people and novel objects. *Applied Animal Behaviour Science.* 45 (1), 109-124.

McDougall, P. Reale, D. Sol, D. Reader, S. (2006). Wildlife conservation and animal temperament: causes and consequences of evolutionary change for captive, reintroduced, and wild populations. *Animal Conservation*. 9 (1), 39-48.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, *22*(3), 276-282.

McKinley, J. Sambrook, T. (2000). Use of human given cues by domestic dogs (*Canis familiaris*) and horses *(Equus caballus). Animal Cognition*. 3(1), 13-22.

Mehrkam, L. R., & Wynne, C. D. (2014). Behavioral differences among breeds of domestic dogs (Canis lupus familiaris): Current status of the science. *Applied Animal Behaviour Science*, *155*, 12-27.

Milgram, N. Head, E. Zicker, S. Candacelkeda, D. Murphy, H. Muggenberg, B. Siwak, C. Tapp, D. Lowry, S. Cotman, C. 2004. Long-term treatment with antioxidants and a program of behavioral enrichment reduces age-dependent impairment in discrimination and reversal learning in beagle dogs. Experimental Gerontology. 39(5), 753-765. https://doi.org/10.1016/j.exger.2004.01.007

Mineka, S. Watson, D. Clark, L. 1998. Comorbidity of anxiety and unipolar mood disorders. *Annual Review Psychology.* 49 (1), 377-412

Mirko, E. Doka, A. Miklosi, A. 2013. Association between subjective rating and behaviour coding and the role of experience in making video assessments on the personality of the domestic dog (Canis familiaris). *Applied Animal Behaviour Science*. 149(1), 45-54.

Mirko, E. Kubinyi, E. Gacsi, M. Miklosi, A. 2012. Preliminary analysis of an adjective-based dog personality questionnaire developed to measure some aspects of personality in the domestic dog (*Canis familiaris*). *Applied Animal Behaviour Science*. 138(1), 88-98.

Mogg, K. Bradley, B. 1998. A cognitive-motivational analysis of anxiety. *Behavioural Research Theories*. 36 (1), 809-848

Mogg, K. Bradley, B. 2005. Attentional bias in generalized anxiety disorder versus depressive disorder. *Cognitive Theoretical Research*. 29 (1), 29-45

Moher, D. Liberati, A. Tetzlaff, J. Altman, D. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine*. 6 (7)

Moore, G. Burkman, K. Carter, M. Peterson M. 2001. Causes of death or reasons for euthanasia in military working dogs: 927 cases (1993-1996). *Scientific Reports: retrospective Study.* 219(2), 209-214.

Momozawa, Y. Ono, T. Sato, F. Kikusui, T. Takeuchi, Y. Mori, Y. Kusonose, R. (2003). Assessment of equine temperament by a questionnaire survey to caretakers and evaluation of its reliability by simultaneous behaviour test. *Applied Animal Behaviour Science*. 84 (1), 127-138.

Morris, T. (2000). Psychological characteristics and talent identification in soccer. *Journal of Sports Sciences*. 18 (9), 715-726.

Mulhoff, N. Stvens, J. Reader, S. (2011). Spatial discounting of food and social rewards in guppies (*Poecilia reticulate*). *Frontiers in Psychology*. 2(68), p1-11

Murphy, J. (1995). Assessment of the temperament of potential guide dogs. *Anthrozoos.*8 (1), 224-228.

Murray, M. Leavenworth, K. 1998. The Contributions of the Military Working Dog in Vietnam. Masters Thesis. Military Operations, Strategy & Tactics. Accession Number: ADA 350072.

Netto, W. Planta, D. (1997). Behavioural testing for aggression in the domestic dog. *Applied Animal Behaviour Science*. 52 (3-4), 243–263.

Neitz, J. Geist, T. Jacobs, G. 1989. Colour Visions in the Dog. *Visual Neuroscience*. 3(2), 119-125. https://doi.org/10.1017/S0952523800004430. Published online: 01 June 2009

Nolan, R. Gravitte, D. (1977). Mine Detecting Canine, 2217.

Nygren, T. Isen, A. Taylor, P. Dulin, J. 1996. The influence of positive affect on the decision rule in risky situations. *Organizational Behavioral Human Decision Processes*. 66 (1), 59-72

Ostaszewski, P. (1996). The relation between temperament and discounting. *European Journal of Personality*. 10 (1), 161-172.

Palestrini, C., Previde, E. P., Spiezio, C., & Verga, M. (2005). Heart rate and behavioural responses of dogs in the Ainsworth's Strange Situation: a pilot study. *Applied Animal Behaviour Science*, *94*(1), 75-88.

Palisin, H. (1986). Preschool Temperament and Performance on Achievement Tests. *Developmental Psychology*. 22 (6), 766-770.

Papale, A. Stott, J. Powell, N. Regier, P. Redish, D. (2012). Interactions between deliberation and delay discounting in rats. *Cognitive Affective Behavioural Neuroscience.* 12(1), p513-526.

Pastore, C. Pirrone, F. Balzarotti, F. Faustini, M. Pierantoni, L. Albertini, M. 2011. Evaluation of physiological and behavioral stress-dependent parameters in agility dogs. *Journal of Veterinary Behaviour: Clinical Applications and Research.* 6(3), 188-194.

Paul, E. Harding, E. Mendl, M. 2005. Measuring emotional processes in animals: the utility of a cognitive approach. *Neuroscience Biobehavioural Reviews*. 29 (1), 469-491.

Peremans, K. Audentart, K. Coopman, F. Hoybergs, Y. Slegers, G. Van Bree, H. Verschooten, F. Dierckx, R. (2002). Functional brain imaging of serotonin-2a receptors in impulsive dogs: A pilot study. *Flemish Veterinary Journal.* 71 (1), 340-347.

Perry, J. Nelson, S. Anderson, M. Morgand, A. Carrol, M. (2007). Impulsivity (delay discounting) for food and cocaine in male and female rats selectively bred for high and low saccharin intake. *Pharmacology, Biochemistry and Behaviour.* 86(1), p822-837

Pongracz, P. Miklosi, A. Vida, V. Csanyi, V. (2005). The pet dog's ability for learning from a human demonstrator in a detour task is independent from the breed and age. *Applied Animal Behaviour Science, 90,* 309–323.

Poulsen, A. Lisle, A. Phillips, C. (2010). An Evaluation of a Behaviour Assessment to Determine the Suitability of Shelter Dogs for Rehoming. *Veterinary Medicine International*. 1 (1), 1-9.

Pritchard, J. Lindberg, A. Main, D. Whay, H. (2005). Assessment of the welfare of working horses, mules and donkeys, using health and behaviour parameters. *Preventive Veterinary Medicine*. 69 (3-4), 265–283.

Puumala, T. Sirvo, J. (1999). Changes in activities of dopamine and serotonin systems in the frontal cortex underlie poor chance accuracy and impulsivity if rats in an attention task. *Neuroscience.* 83(2), p489-499

Reale, D. Reader, S. Sol, D. McDougall, P. Dingemanse, N. (2007). Integrating animal temperament within ecology and evolution. *Biological Reviews*. 82 (1), 291-318.

Reale, D. Gallant, B. Leblanc, M. Festa-Bianchet, M. (2000). Consistence of temperament in bighorn ewes and correlates with behaviour and life history. *Animal Behaviour*. 60 (1), 589-597.

Reimer, S. Mills, D. Wright, H. (2014). Impulsive for life? The nature of long –term impulsivity in domestic dogs. *Animal Cognition*. 17, 815-819.

Reynolds, B. Wit, H. Richards, J. (2002). Delay of gratification and delay discounting in rats. *Behavioural Processes*. 59 (1), 157-168.

Robazza, C. Gallina, S. D'Amico, M. Izzicupo, P. Bascelli, A. Di Fonso, A. Mazzaufo, C. Capobianco, A. Di Baldassare, A. (2012). Relationship between biological markers and psychological states in elite basketball players across a competitive season. *Psychology of Sport and Exercise*. 13 (1), p509-517.

Robazza, C. Pellizzari, M. Hanin, Y. (2004). Emotion self-regulation and athletic performance: An application of the IZOF model. *Psychology of Sport and Exercise*. 5 (1), p379-404.

Roberts, B. Kuncel, N. Shiner, R. Caspi, A. Goldberg, L. (2007). The Power of Personality: The Comparative Validity of Personality Traits, Socioeconomic Status, and Cognitive Ability for Predicting Importnat Life Outcomes. *Perspectives on Psychological Science*. 2 (4), 313-345.

Rooney, N., Bradshaw, J., Almey, H., Attributes of Specialist Search Dogs—A Questionnaire Survey of UK Dog Handlers and Trainers, *Journal of Forensic Sciences*, Vol. 49, No. 2, 2004, pp. 1-7, https://doi.org/10.1520/JFS2003048. ISSN 0022-1198

Rooney, N. Cowan, S. 2011. Training methods and owner–dog interactions: Links with dog behaviour and learning ability. *Applied Animal Behaviour Science*. 132(3-4), 169-177. https://doi.org/10.1016/j.applanim.2011.03.007

Rooney, N. J., Gaines, S. A., Bradshaw, J. W. S., & Penman, S. (2007). Validation of a method for assessing the ability of trainee specialist search dogs. *Applied Animal Behaviour Science*, *103*, 90–104.

Rosati, A. Stevens, J. Hare, B. Hauser, M. (2007). The evolutionary origins of human patience: temporal preferences in chimpanzees, bonobos, and human adults. *Current Biology*. 17 (19), 1663–1668.

Romeyer, A. Bouissou, M. (1992). Assessment of fear reactions in domestic sheep, and influence of breed and rearing conditions. *Applied Animal Behaviour Science.* 34(1), 93-119.

Rothbart, M. Ahadi, S. Hershey, K. (1994). Temperament and Social Behaviour in Childhood. *Merrill-Palmer Quarterly*. 40 (1), 21-39.

Rothbart, M. Evans, D. Ahadi, S. (2000). Temperament and Personality: Origins and Outcomes. *Journal of Personality and Social Psychology*. 78 (1), 122-135.

Ruefenacht, S. Gebhardt-Henrich, S. Miyake, T. Gaillard, C. (2002). A behaviour test on German Shepherd dogs: heritability of seven different traits. *Applied Animal Behaviour Science*. 79(2), 113-132.

Ruiz, M. Hanin, Y. Robazza, C. (2015). Multimodal assessment of athletes' performance-related feeling states: Developing the psychobiosocial states (PBS-S) scale. Submitted.

Ruiz, M. Raglinband, J. Yuri L. (2015). The individual zones of optimal functioning (IZOF) model (1978–2014): Historical overview of its development and use, accessed on: https://www.researchgate.net/publication/276831504_The_individual_zones_of_optimal_functioning_IZOF_model_1978-2014_Historical_overview_of_its_development_and_use [accessed Mar 29, 2017].

Rust, J. Golombok, S. Kosinski, M. Stillwell, D. (2014) Modern Psychometrics, Third Edition: The Science of Psychological Assessment. 3rd Edition. ISBN 978-0-415-44216-9. Routeledge. Taylor & Francis Group. Psychology press

Russell, J. (2009) Emotion, core affect, and psychological construction. *Cognition and Emotion*. 23 (7), 1259-1283

Salminen, S. Liukkonen, J. Hanin, Y. Hyvonen, A. (1995). Anxiety and athletic performance of Finnish athletes: Application of the Zone of Optimal Functioning Model. *Personality and Individual Differences*. 19 (5), p725-729.

Sanders, C. R. (2000). The impact of guide dogs on the identity of people with visual impairments. *Anthrozoös*, *13*(3), 131-139.

Sanson, A. Hemphill, S. Smart, D. (2004). Connections between Temperament and Social Development: A Review. *Social Development*. 13 (1), 142-170.

Schoberl, I. Kortekaas, K. Schoberl, F. Kotrschal. 2015. Algorithm-supported visual error correction (AVEC) of heart rate measurements in dogs, Canis lupus familiaris. *Behavior Research Methods.* 47(4), 1356-1364

Scott, J. P. & Fuller, J. L. (1965). *Genetics and the Social Behaviour of the Dog*. University of Chicago Press, Chicago, p. 468.

Seaman, S. Davidson, H. Waran, N. 2002. How reliable is temperament assessment in the domestic horse (*Equus callabus*). *Applied Animal Behaviour Science*. 78(1), 175-191.

Seibert, S. Crant, M. Kraimer, M. (1999). Proactive Personality and Career Success. *Journal of Applied Psychology*. 84 (3), 416-427.

Seibert, S. Kraimer, M. (2001). The Five-Factor Model of Personality and Career Success. *Journal of Vocational Behaviour*. 58 (1), 1-21.

Serpell, J. Hsu, Y. (2005). Effects of breed, sex, and neuter status on trainability in dogs. *Anthrozoos*. 18(3), 196-207.

Serpell, J. Hsu, Y. (2001). Development and validation of a novel method for evaluating a behaviour and temperament in guide dogs. *Applied Animal Behaviour Science.* 72 (1), 347-364.

Shafir, S. Reich, T. Tsur, E. Erev, I. Lotem, A. 2008. Perceptual accuracy and conflicting effects of certainty on risk-taking behaviour. *Nature*. 453 (1), 917-951

Sheppard, G .Mills, D. 2003. Construct Models in Veterinary Behavioural Medicine: Lessons from the Human Experience. *Veterinary Research Communications.* 27(1), 175-191.

Sheppard, G. Mills, D. (2002). The development of a psychometric scale for the evaluation of the emotional predispositions of pet dogs. *International Journal of Comparative Psychology*. 15 (1), 201-222

Sherman, B. L., Gruen, M. E., Case, B. C., Foster, M. L., Fish, R. E., Lazarowski, L., et al. & Dorman, D. C. (2015). A test for the evaluation of emotional reactivity in Labrador retrievers used for explosives detection. *Journal of Veterinary Behavior: Clinical Applications and Research*, *10*(2), 94-102.

Sih, A., Bell, A. M., Johnson, J. C., & Ziemba, R. E. (2004). Behavioral syndromes: An integrative overview. *Quarterly Review of Biology*, **79**, 241–277.

Sinn, D. Gosling, S. Hilliard, S. (2010). Personality and performance in military working dogs: Reliability and predictive validity of behavioural tests. *Applied Animal Behaviour Science*. 127 (1), p51-65.

Smith, J. Shields, W. Washburn, D. 2003. The comparative psychology of uncertainty monitoring and metacognition. *Behavioral Brain Science*. 26 (1), 317-339

Southwick, S. Vythilingam, M. Charney, D. (2005). The Psychobiology of Depression and Resilience to Stress: Implications for Prevention and Treatment. *Annual Review of Clinical Psychology***.** DOI: 10.1146/annurev.clinpsy.1.102803.143948. 1 (1), 255-291.

Spedding, C.R.W. (1995). Sustainability in animal production systems. *Animal Science, 61,* 1–8.
Slabbert, J. Odendaal, J. (1999). Early prediction of adult police dog efficiency—a longitudinal study. *Applied Animal Behaviour Science*. 64 (4), 269–288.

Stamps, J. Briffa, M. Brio, P. (2012). Unpredictable animals: individual differences in intraindividual variability (IIV). *Animal Behaviour.* 83(1), 1325-1334.

Stamps, J. Groothuis, T. (2010). The development of animal personality: relevance, concepts and perspectives. *Biological Reviews*. 82 (2), 301-325.

Stevens, JR. Rosati, AG. Ross, KR. Hauser, MD. (2005). Will Travel for Food: Spatial Discounting in Two New World Monkeys. *Current Biology*. 15 (20), 1855–1860

Sutcliffe, J. Marshall, K. Neill, J. (2007). Influence of gender on working and spatial memory in the novel object recognition task in the rat. *Behavioural Brain Research.* 177(1), 117-125.

Svartberg, K. (2005). A comparison of behaviour in test and in everyday life: evidence of three consistent boldness-related personality traits in dogs. *Applied Animal Behaviour Science*. 91 (1), p103-128

Svartberg, K. (2003). Shyness-boldness predicts performance in working dogs. *Applied Animal Behaviour Science*. 79 (1), p157-174

Svartberg, K. Forkman, B. (2002). Personality traits in the domestic dog (Canis familiaris). *Applied Animal Behaviour Science*. 79 (1), p133-155

Svartberg, K. Tapper, I. Temrin, H. Radesater, T. Thorman, S. (2005). Consistency of personality traits in dogs. *Animal Behaviour*. 69 (1), p283-291.

Svobodová, I., Vápeník, P., Pinc, L., and Bartoš, L. (2008) Testing German shepherd puppies to assess their chances of certification*. Applied Animal Behaviour Science,* 113(1–3), pp. 139-149.

Szetei, V. Miklosi, A. Topal, J. Csanyi, V. (2003). When dogs seem to lose their nose: an investigation on the use of visual and olfactory cues in communicative context between dog and owner. *Applied Animal Behaivour science.* 83(1), 141-152.

Tamimi, N. Jamshidi, S. Serpell, J.A. Mousavi, S. Ghasempourabadi, Z. 2015. Assessment of the C-BARQ for evaluating dog behaviour in Iran. *Journal of Veterinary Behaviour: Clinical Applied Research*. 10(1), 36-40.

Taylor, K. Mills, D. (2006). The development and assessment of temperament tests for adult companion dogs. *Journal of Veterinary Behaviour: Clinical Applications and Research*. 1 (3), 94–108.

Temesi, A. Turcsan, B. Miklosi, A. 2014. Measuring fear in dogs by questionnaires: an exploratory study toward a standardized inventory. *Applied Animal Behaviour Science*. 161(1), 121-130.

Thompson, J. (2005). Proactive Personality and Job Performance: A Social Capital Perspective. *Journal of Applied Psychology*. 90 (5), 1011-1017.

Tobin, H. Logue, A. Chelonis, J. Ackerman, K. (1996). Self-control in the monkey *Macaca fascicularis. Animal Learning and Behaviour.* 24, p168-174

Tomkins, L. M., Thomson, P. C., & McGreevy, P. D. (2011). Behavioral and physiological predictors of guide dog success. *Journal of Veterinary Behavior: Clinical Applications and Research*, *6*(3), 178-187.

Tschirky, F. Brabec, B. Kern, M. (2000). Avalanche rescue systems in Switzerland: experience and limitations, in Proceedings at the International Snow Science Workshop (ISSW 2000), USA. 369-376.

Trybocka, R. (2010). Character assessment testing to test suitability for guide dogs. *Veterinary Nursing Journal*. DOI: 10.1111/j.2045-0648.2010.tb00101.

Turcsan, B. Kubinyi, E. Miklosi, A. (2011). Trainability and boldness traits differ between dog breed clusters based on conventional breed categories and genetic relatedness. *Applied Animal Behaviour Science.* 132(1), 60-71.

Van der Borg, J. Netto, W. Planta, D. (1991). Behavioural testing of dogs in animals shelters to predict problem behaviour. *Applied Animal Behaviour Science.* 32 (1), 237-251.

Verbeek, M. Boon, A. Drent, P. (1996). Exploration, aggressive behaviour and dominance in pair-wise confrontations of juvenile male great tits. *Behaviour.* 133(1), 945-963.

Visser, E. Van Reenen, C. van der Werf, J. Schilder, M. Knaap, J. Barnevelde, A. Blokhuis, H. (2002). Heart rate and heart rate variability during a novel object test and a handling test in young horses. *Physiology and Behaivour.* 76(1), 289-296.

Walther, S., Yamamoto, M., Thigpen, A. P., Garcia, A., Willits, N. H., & Hart, L. A. (2017). Assistance dogs: historic Patterns and roles of dogs placed by aDi or igDF accredited facilities and by non-accredited facilities. *Frontiers in Veterinary Science*, *4*. https://doi.org/10.3389/fvets.2017.00001.

Wasser, S. Davenport, B. Ramage, E. Hunt, K. Parker, M. Clarke, C. Stenhouse, G. (2004). Scent detection dogs in wildlife research and management: application to grizzly and black bears in the yellowhead ecosystem, Alberta, Canada. *Canadian Journal of Zoology.* 82(1), 475-492.

Weiner, P. Haskell, M. 2016. Use of questionnaire-based data to assess dog personality. *Journal of Veterinary Behaviour*. 16(1), 81-85.

Wilsson, E. Sundgren, P. (1997). The use of a behaviour test for the selection of dog for service and breeding, I: Method of testing and evaluating test results in the adult dog, demands on different kinds of service dogs, sex and breed differences. *Applied Animal Behaviour Science*. 53 (1), 279-295.

Wilsson, E. and Sundgren, P. (1998) Behaviour test for eight-week old puppies—heritability's of tested behaviour traits and its correspondence to later behaviour. *Applied Animal Behaviour Science,* 58(1–2), pp. 151-162.

Wilsson, E. Sinn, D. (2012). Are there differences between behavioral measurement methods? A comparison of the predictive validity of two ratings methods in a working dog program. *Applied Animal Behaviour Science*. 141 (3-4), 158–172.

Winston, J. Gottfried, J. Kilner, J. Dolan, R. 2005. Integrated neural representations of odour intensity and affective valence in human amygdala. *Journal of Neuroscience.* 25(39), 8903–8907. [PubMed: 16192380]

Winston, J. O'Doherty, J. Dolan, R. 2003. Common and distinct neural responses during direct and incidental processing of multiple facial emotions. *Neuroimaging.* 20(1), 84–97. [PubMed: 14527572]

Wittenborn, J. S., Zhang, X., Feagan, C. W., Crouse, W. L., Shrestha, S., Kemper, A. R., et al. & Vision Cost-Effectiveness Study Group. (2013). The economic burden of vision loss and eye disorders among the United States population younger than 40 years. *Ophthalmology*, *120*(9), 1728-1735.

Wolff, M. Leander, D. (2002). Selective Serotonin Reuptake Inhibitors Decrease Impulsive Behaviour as Measured by an Adjusting Delay Procedure in the Pigeon. *Neuropsychopharmacology*. 27(3), p421-429.

Wolf, M., van Doorn, G. S., Leimar, O., & Weissing, F. J. (2007). Life-history trade-offs favour the evolution of animal personalities. *Nature*, **447**, 581–584.

Wright, W. Bower, G. 1992. Mood effects on subjective probability assessment. *Organizational Behavioral Human Decision Processes.* 52 (1), 276-291

Wright, H. Mills, D. Pollux, P. (2012). Behavioural and physiological correlates of impulsivity in the domestic dog (*Canis familiaris*). *Physiology and Behaviour*. 105(3), 676-682.

Wright, H. Mills, D. Pollux, P. (2011). Development and Validation of a Psychometric Tool for Assessing Impulsivity in the Domestic Dog (*Canis familiaris*). *International Journal of Comparative Psychology*. 24(1), p210-225.

Yen, M., & Lo, L. H. (2002). Examining test-retest reliability: an intra-class correlation approach. *Nursing Research*, *51*(1), 59-62.

Yerkes, R. M. & Dodson, J. D. (1908). The Relationship of Strength of Stimulus to Rapidity of Habit Formation. *Journal of Comparative Neurology and Psychology*. 1908, 18, 459-482.

Zubedat, S. Aga-Mizrachi, S. Cymerblit-Sabba, A. Shwartz, J. Leon, J. Rozen, S. Varkovitzky, I. Eshed, Y. Grinstein, D. Avital, A. (2014). Human-animal interface: The effects of handler's stress on the performance of canines in an explosive detection task. *Applied Animal Behaviour Science.* 158(1), 69-75.

Appendices

Appendix 1: Scent Work Study

A.) Search Plans.

	search number									
Box location	1	2	3	4	5	6	7	8	9	10
1	1		B			U		1		E1
2	B	D1		1	1	D	U	E		
3			U1	E		E1			D	1
4	1	B			E1	U	1	D		
5	E		1			D		1		1
6		D		1		E			1	
7	D	1		B			1			U
8			1				U	1	D	E
9	E1			U	1	1		B		D
10			E		E			1	U	1
11		U		D1		1		E	1	
12			U		D		1			1
13	U	1		D	1				E	
14		E	1	1				U		B
15	1		D				B		1	U
16	U				D		1		1	
17		U	E	U	1	B	D1			
18		1			U	1	E			D
19	1	E	D	1	U					
20			1		B		D	U1	E	
21	D	1		E		1			U1	

In each search one article will be placed in a harder location-either in an elevated, difficult or upright box.

1= article
b= blank-no hide box to search
e= elevated hide
d= box part closed (difficult)
u= box upright
all other boxes on side with opening facing away from search entry point

B. **Hand out given to all owners to guide them through the search process**
Pam Mackinnon, Talking Dogs Scentwork® 2015

HANDLING TIPS

Holding the lead

Make sure that you get the lead out of your hand while searching, you will need both hands to search and the lead will get in the way.

Following the dog

Ask yourself how following will help the search. Instead, stay in front of the dog to keep the search flowing and providing him with lots of places to look next.

Not asking the question

The question helps the dog re-examine the area, and so allows him to say no, it's not the scent he's looking for, or yes, that smells good. It can help strengthen the indication and supports the dog if he's unsure of what he's smelling.

Not moving away when asking the question

If you stay where you are or, even worse, move towards the dog when asking the question, you can skew the answer. The dog must be able to reply yes or no to the question. If you stay where you are you don't give him the obvious opportunity to move away from the incorrect scent. If you move towards it, there's a great danger that he will think you know the answer is yes. Very confusing if the find is not there. And if it is there, the dog can stop searching as well, safe in the knowledge that you are doing the searching!

Leaving the dog to do all the work

When your dog is on the scent but is having trouble finding the source, instead of throwing your hands up and just asking him to keep searching, move forward to suggest areas or items to search. Leaving him to think up new places to look, without support from you, is inefficient and can be stressful for your dog.

Not covering the whole area / getting stuck in one spot

It is very easy to get stuck in one part of the search area and so neglect other parts. This often happens if you are following the dog. Keep moving so that you cover the whole area, giving the dog access to everything. This becomes less of a problem once you have a search plan, i.e. learnt the search patterns.

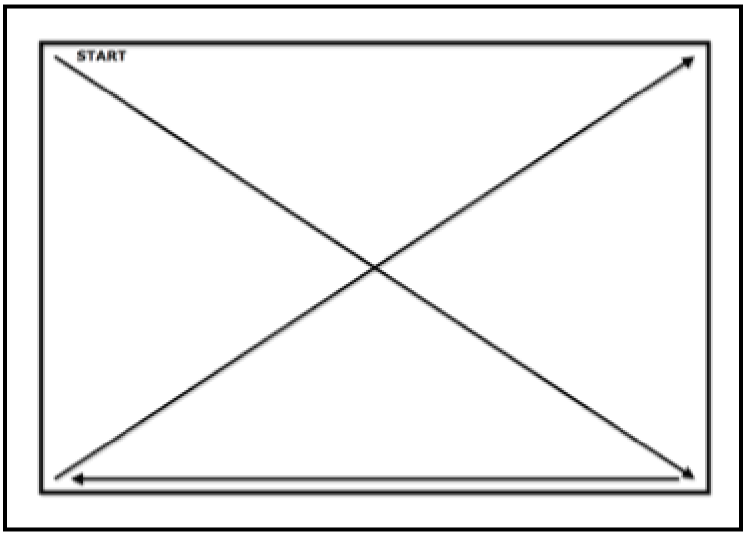
DIRECTED SEARCH

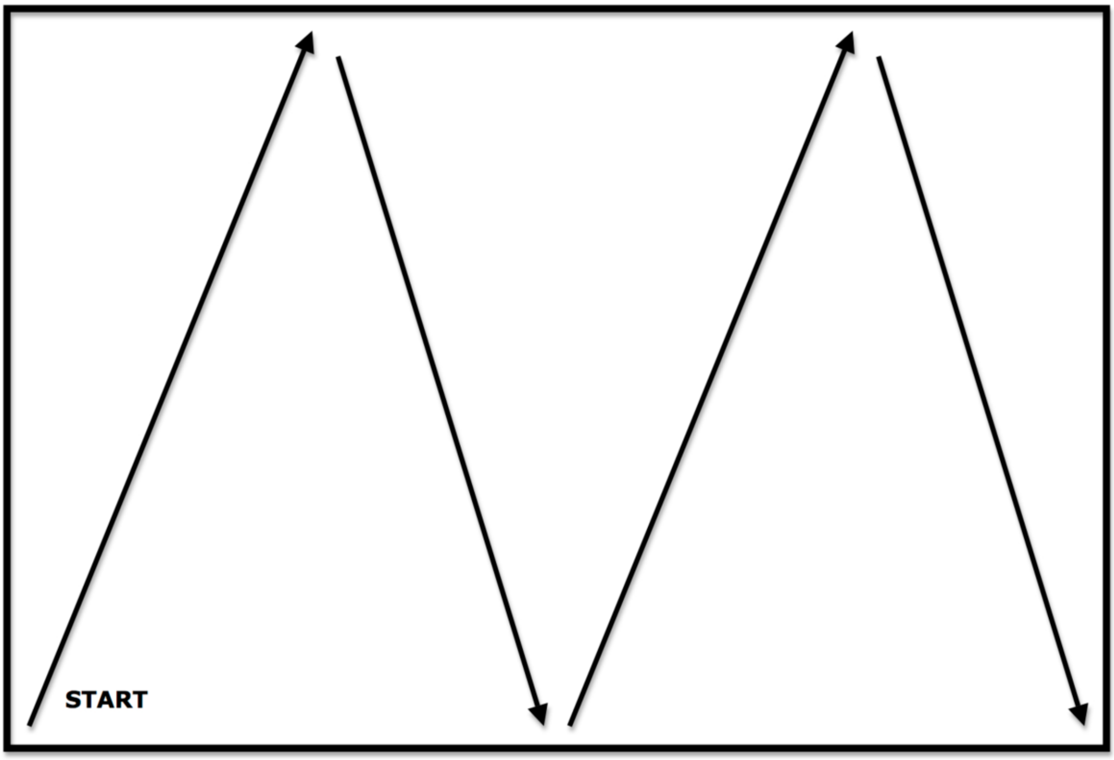
The **directed search** is a much more specific and thorough search. The purpose of this type of search is to help the dog examine the entire area ensuring he checks high, low and everything in between. This technique will form the main part of the majority of the searches you conduct from now on.

SEARCH PLAN

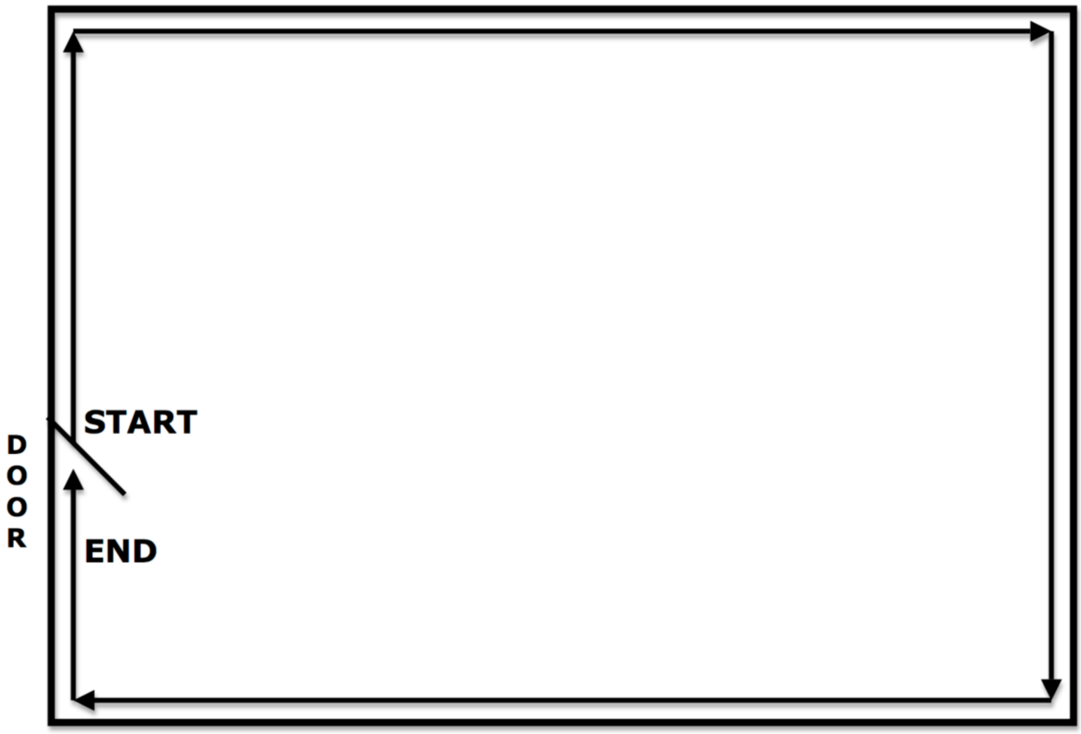
♦**Free** – first chance for the dog to settle, get used to the area and find the article.

♦**Corners or zigzags** - depending on the size/shape of the area.

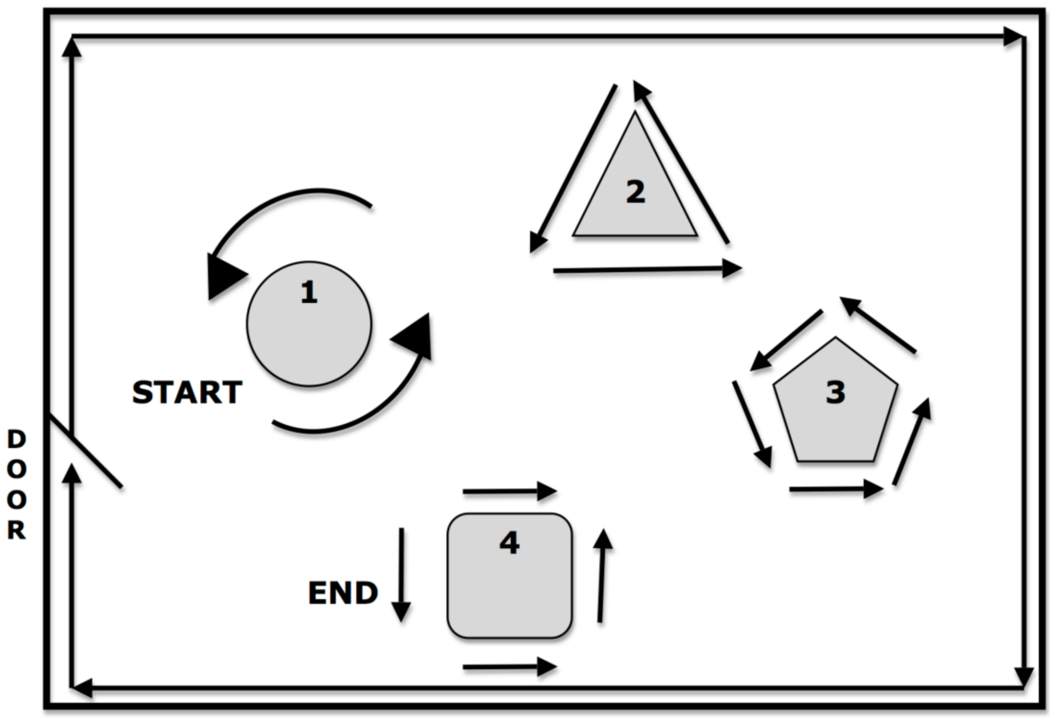




♦**Perimeter** – start to the right of the door where you entered the search area.



♦**Interior** – work your way around all the items inside the search area.



C. Order of conditions for searches, half of the population followed set 1 in order, the other half followed set 2.

Set 1	Order of conditions
1	Obedience exercises (ask dog to perform a sit and a down)
2	Get your dog moving over a series of small jumps (as quickly as they are able to)
3	Suddenly switch you attention from talking to the researcher to working your dog
4	Play with your dog with a toy-maximum excitement
5	Settle on bed 10 minutes
6	Tease dog with toy-dog does not get to access toy
7	Neutral release of dog to search
8	Are you ready? Gear dog up to work (moderate excitement)
9	Food bowl taken away while eating (this can be done by the owner if preferred)
10	Walk your dog at a steady pace around the waiting room

Set 2	Order of conditions
1	Walk your dog at a steady pace around the waiting room
2	Food bowl taken away while eating (this can be done by the owner if preferred)
3	Are you ready? Gear dog up to work (moderate excitement)
4	Neutral release of dog to search
5	Tease dog with toy-dog does not get to access toy
6	Settle on bed 10 minutes
7	Play with your dog with a toy-maximum excitement
8	Suddenly switch you attention from talking to the researcher to working your dog
9	Get your dog moving over a series of small jumps (as quickly as they are able to)
10	Obedience exercises (ask dog to perform a sit and a down)

Appendix 2 Systematic Review

Papers Selected for Review:

Asher, L. Blythe, S. Roberts, R. Toothill, L. Craigon, P. Evans, K. Green, M. Englang, G. (2013). A standardised behaviour test for potential guide dog puppies: Methods and association with subsequent success in guide dog training. *Journal of Veterinary Behaviour*. 8 (1), p431-438

Batt, L. Batt, M. Baguley, J. McGreevy, P. (2008). Factors associated with success in guide dog training. *Journal of Veterinary Behaviour*. 3 (1), p143-151.

De Meester, R. De Bacquwr, D. Peremans, K. Vermeire, S. Planta, D. Coopman, F. Audenaert, K. (2008). A preliminary study on the use of the Socially Acceptable Behaviour test as a test for shyness/confidence in the temperament of dogs. *Journal of Veterinary Behaviour*. 3 (1), p161-170.

De Meester, R. Pluijmakers, J. Vermeire, S. Laevens, H. (2011). The use of the Socially Acceptable Behaivour test in the study of temperament of dogs. *Journal of Veterinary Behaviour*. 6 (1), p211-224.

Fuchs, T. Gaillard. C, Gebhardt-Henrich, S. Ruefenacht, S. Steiger, A. (2005). External factors and reproducibility of the behaviour test in German shepherd dogs in Switzerland. *Applied Animal Behaviour Science*. 94 (1), p287-301.

Sinn, D. Gosling, S. Hilliard, S. (2010). Personality and performance in military working dogs: Reliability and predictive validity of behavioural tests. *Applied Animal Behaviour Science*. 127 (1), p51-65.

Slabbert, J. Odendaal, J. (1999). Early prediction of adult police dog efficacy-a longitudinal study. *Applied Animal Behaviour Science*. 64 (1), p269-288.

Svartberg, K. (2005). A comparison of behaviour in test and in everyday life: evidence of three consistent boldness-related personality traits in dogs. *Applied Animal Behaviour Science*. 91 (1), p103-128

Svartberg, K. (2002). Shyness-boldness predicts performance in working dogs. *Applied Animal Behaviour Science*. 79 (1), p157-174

Svartberg, K. Forkman, B. (2002). Personality traits in the domestic dog (Canis familiaris). *Applied Animal Behaviour Science*. 79 (1), p133-155

Svartberg, K. Tapper, I. Temrin, H. Radesater, T. Thorman, S. (2005). Consistency of personality traits in dogs. *Animal Behaviour*. 69 (1), p283-291.

Svobodova, I. Vapenik, P. Pinc, L. Bartos, L. (2008). Testing German shepherd puppies to assess their chances of certification. *Applied Animal Behaviour Science*. 113 (1), p139-149

Wilsson, E. Sundgren, P. (1998). Behaviour test for eight week old puppies-heritability's of tested behaviour traits and its correspondence to later behaviour. *Applied Animal Behaviour Science*. 58 (1), p151-162

Appendix 3: Working Dog Questionnaire

UNIVERSITY OF LINCOLN

Handler Name (optional):

Email/preferred contact (optional):

Dog's name: Dog's breed:

Dog's Age (or age withdrawn): Dog's sex:

Is your dog neutered?

Where does your dog usually live (please circle):

At home with you | at work in kennels | elsewhere

Role of dog (please circle):

Detection | Patrol | other (please specify):

Working Status (please circle):

Active | Retired | Training | Withdrawn

Reason for withdrawal:

How successful is/was your dog when working in real life situations? (5 being successful almost 100% of the time, 3 being 50%, 1 being almost never). Please circle

1 | 2 | 3 | 4 | 5 | N/A-still in training

How successful is/was your dog in training? Please circle

1 | 2 | 3 | 4 | 5

Source:

Purpose bred | Rescue | Private Dealer

Other: please specify if known

Positive and negative activation scale for dogs

For each of the statements below, please place a cross in the box which most accurately describes your level of agreement with how your dog behaves in general in this situation. Please consider whether your dog's behaviour is of similar intensity and occurs as frequently as described. For example, if in item 2 you dog **always** becomes **a little** excited when it is about to go for a walk, you would mainly agree with the statement.

If your dog has never encountered the situation and you are unable to predict the behaviour, please use the not applicable option

		Agree strongly	Mainly agree	Partly agree, partly disagree	Mainly disagree	Disagree strongly	Not applicable	Office use only
1	Your dog is rarely frightened							
2	Your dog becomes very excited when it is about to go for a walk (e.g. when it sees its lead, or when it hears "walkies", etc.)							
3	Your dog is easily startled by noises and / or movements							
4	Your dog is very persistent in its efforts to get you to play							
5	Your dog shows little interest in its surroundings							
6	Your dog appears nervous and / or jumpy for several minutes after it has been startled							
7	Your dog is easily excited							
8	Your dog has a specific fear or phobia							
9	Your dog tries to escape from the garden							
10	Your dog appears calm in noisy, crowded places							
11	Your dog is full of energy							
12	Your dog is frightened by noises from the television or radio							

		Agree strongly	Mainly agree	Partly agree, partly disagree	Mainly disagree	Disagree strongly	Not applicable	Office use only
13	Your dog usually appears relaxed							
14	Your dog is lazy							
15	Your dog adapts quickly to changes in its environment (e.g. being cared for by different people, moving house or a family member leaving home)							
16	Your dog appears afraid of the vacuum cleaner or any other familiar household appliance							
17	Your dog requires a great deal of encouragement to take part in energetic activities							
18	Your dog persists in being naughty despite being told off for the behaviour							
19	Your dog appears calm in unfamiliar environments							
20	Your dog is very boisterous							
21	Your dog appears unsettled by changes to its routine (e.g. if it is not fed at the usual time, if it is left alone for longer than usual)							

DIAS:

For each of the statements below please place a cross in the box that most accurately describes your level of agreement: The answer should reflect the *general personality of the dog*, so for example if a statement applies to your dog in some situations but not others, please make a judgement as to how much you agree

		Strongly agree	Mainly agree	Partly agree, partly disagree	Mainly disagree	Strongly disagree	Don't know / not applicable
1	My dog shows extreme physical signs when excited (e.g. drooling, panting, raising hackles, urination, licking lips, widening of eyes)						
2	When my dog gets very excited it can lead to fixed repetitive behaviour (i.e. an action that is repeated in the same way over and over again), such as tail chasing or spinning around in circles						
3	I would consider my dog to be very impulsive (i.e. has sudden, strong urges to act; acts without forethought; acts without considering effects of actions)						
4	My dog doesn't like to be approached or hugged						
5	My dog becomes aggressive (e.g. growl, snarl, snap, bite) when excited						
6	My dog appears to be 'sorry' after it has done something wrong						
7	My dog does not think before it acts (e.g. would steal food without first looking to see if someone is watching)						
8	My dog can be very persistent (e.g. will continue to do something even if it knows it will get punished or told off)						
9	My dog may become aggressive (e.g. growl, snarl, snap, bite) if frustrated with something						

10	My dog is easy to train						
		Strongly agree	Mainly agree	Partly agree, partly disagree	Mainly disagree	Strongly disagree	Don't know / not applicable
11	My dog is not keen to go into new situations						
12	My dog takes a long time to lose interest in new things						
13	My dog calms down very quickly after being excited						
14	My dog appears to have a lot of control over how it responds						
15	My dog is very interested in new things and new places						
16	My dog reacts very quickly						
17	My dog is not very patient (e.g. gets agitated waiting for its food, or waiting to go out for a walk)						
18	My dog seems to get excited for no reason						

Thank you very much for taking part.

Appendix 4: Behaviour Tests

Results from Behavioural Tests

Test Data	Attention Test		Discrimination Test		Social Cues Test		Novel Object Test-Block Interaction		Novel Environment Test		Bin Interaction	Social Sensitivity Test	
	Opt.	Sub.	Opt.	Sub.	Opt.	Sub.	Opt.	Sub.	Opt.	Sub.		Opt.	Sub.
Correct	28.1 +/- 2.1	31.8 +/- 5	36.4 +/- 5	21.3 +/- 5	2.5 +/- 0.6	1.6 +/- 0.3	Block Sniffing time 1.4 +/- 0.4	Block Sniffing time 1.72 +/- 0.7	8.6 +/- 1.4	7.5 +/- 2.2	Latency to interact before voice	3.9 +/- 1.9	6.1 +/-1.9
.p=	0.475		0.051		0.314		0.034	0.14	0.426			0.282	
Incorrect	8 +/- 1.8	7.6 +/- 1.4	0.4 +/- 0.2	1.1 +/- 0.5	11.5 +/-0.9	9.9 +/- 1.5	Bin sniffing time 0.15 +/- 0.24	Bin sniffing time 0.38 +/- 0.24			Interaction before voice	8.9 +/- 2.7	6 +/-2.5
.p=	0.863		0.37		0.35							0.512	
Latency to switch or interact	64.7 +/- 14.6	44.4 +/- 11.8	1.3 +/- 0.8	21.6 +/- 11.7							Startle	1.5 +/- 0.5	2.5 +/-1.4
.p=	0.291		0.074									0.557	
Opt out	5.6 +/- 1.5	5.6 +/- 1.4			3.4 +/-0.6	4.9 +/- 0.8					Door directed behaviour	7.3 +/- 3.8	5.4 +/-1.3
.p=	0.970				0.129							0.349	
Investigation									25.3 +/- 8.0	14.9 +/- 5.0	Latency to interact after	8.4 +/- 2.9	10.4 +/-3.2
.p=									0.314			0.605	
Trials to learn or total time	317.4 +/- 51.5	256 +/- 50.3	46.5 +/- 8.9	100.7 +/- 14.7	321.2 +/- 28.5	304.7 +/- 29.6			118.8 +/- 13.9	102.1 +/- 13.6	Interaction after	15.1 +/- 2.5	15.4 +/-6.4
.p=	0.414		0.021		0.694				0.404			0.282	

Analysis of significant differences between group 1 (Optimal Group) and group 2 (lower scoring Group) during test.

Retest data	Attention Test		Discrimination Test		Social Cues Test		Novel Object Test-Block Interaction		Novel Environment Test		Bin Interaction	Social Sensitivity Test	
	Opt.	Sub.	Opt.	Sub.	Opt.	Sub.	Opt.	Sub.	Opt.	Sub.		Opt.	Sub.
Correct	24.2 +/- 2.0	27.2 +/- 3.6	43.8 +/- 3.6	25.7 +/- 4.3	1.4 +/- 0.7	1.9 +/-0.9	Block Sniffing time 2.42 +/- 0.72	Block Sniffing time 1.38 +/- 0.8	5.9 +/- 1.7	3.2 +/- 1.1	Latency to interact before voice	2.2 +/- 0.7	6.5 +/- 2.6
.p=	0.447		0.007		0.73		0.05	0.225	0.207			0.143	
Incorrect	4.9 +/- 1.3	6.8 +/- 1.3	0.22 +/- 0.15	1 +/- 0.26	11.7 +/- 1.95	9.9 +/-1.4	Bin sniffing time 0.46 +/- 0.26	Bin sniffing time 0.36 +/- 0.13			Interaction before voice	8.3 +/- 3.5	4.8 +/- 2.1
.p=	0.315		0.043		0.475							0.436	
Latency to switch or interact	35.6 +/- 9.8	31.3 +/-5	0.47 +/- 0.3	8.1 +/- 5.2							Startle	1.4 +/- 0.6	0.6 +/- 0.3
.p=	0.697		0.035									0.481	
Opt out	2.8 +/- 0.89	2 +/- 0.68			4.1 +/- 1.1	5.3 +/-1.2					Door directed behaviour	5.1 +/- 1.7	8.4 +/- 2.6
.p=	0.486				0.481							0.308	
Investigation									16.5 +/- 5.9	5.9 +/- 2.1	Latency to interact after	9.3 +/- 3.4	9.5 +/-3
.p=									0.123			0.631	
Trials to learn or total time	267.9 +/- 74.2	222.2 +/- 41	28.3 +/- 2.2	31.8 +/- 2.87	277.2 +/- 30.6	279.9 +/-28.3			112.8 +/- 11.3	89.7 +/- 8.9	Interaction after	12.5 +/- 2.5	12.1 +/- 3.2
.p=	1		0.46		1				0.126			0.923	

Retest Analysis of significant differences between group 1 (Higher Scoring Group) and group 2 (Lower Scoring Group).

Population Test Retest Analysis	Attention Test		Discrimination Test		Social Cues Test		Novel Object Test-Block Interaction		Novel Environment Test		Bin Interaction	Social Sensitivity Test	
	Test	Retest	Test	Retest	Test	Retest	Test	Retest	Test	Retest		Test	Retest
Correct	29.72 +/- 10.38	25.79 +/- 7.72	28.88 +/- 16.69	32.71 +/- 14.8	2.1 +/- 1.61	1.67 +/- 2.28	Block Sniffing time 1.54 +/- 1.73	Block Sniffing time 1.9 +/- 2.41	8.1 +/- 5.77	4.56 +/- 4.7	Latency to interact before voice	4.92 +/- 6.25	4.34 +/- 6.35
Correlation	+0.428 (0.086)		+0.456 (0.066)		+0.122 (0.628)		+0.534 (*0.015)		+0.318 (0.172)			-0.129 (0.588)	
.p=	0.014		0.351		0.320		0.407		0.011			0.563	
Incorrect	7.8 +/- 4.98	5.9 +/- 4.04	0.76 +/- 1.15	0.65 +/- 0.79	10.76 +/- 3.92	10.78 +/- 5.08	Bin sniffing time 0.26 +/- 0.59	Bin sniffing time 0.41 +/- 0.64			Interaction before voice	7.52 +/- 8.38	6.55 +/-9.1
Correlation	-0.027 (0.911)		+0.457 (0.065)		+0.000 (0.999)		+0.080 (0.737)					-0.230 (0.329)	
.p=	0.195		0.623		0.820		0.540					0.658	
Latency to switch or interact	54.55 +/- 42.10	33.35 +/- 22.72	10.76 +/- 25.67	8 +/- 16.15							Startle	2 +/- 3.26	0.98 +/- 1.55
Correlation	-0.363 (0.127)		+0.616 (**0.005)									+0.082 (0.731)	
.p=	0.126		0.780									0.232	
Opt out	5.4 +/- 4.71	2.31 +/- 2.52			4.1 +/- 2.3	4.72 +/- 3.54					Door directed behaviour	6.36 +/- 9.36	6.76 +/- 7.09
Correlation	-0.359 (0.131)				-0.143 (0.572)							-0.032 (0.894)	
.p=	0.036				0.422							0.695	
Investigation									20.46 +/- 22.26	11.21 +/- 14.67	Latency to interact after	6.36 +/- 9.36	6.76 +/- 7.09
Correlation									-0.050 (0.717)			-0.280 (0.233)	
.p=									0.083			0.872	
Trials to learn or total time	290.13 +/- 152.85	243.83 +/- 176.02	70.41 +/- 42.16	32.82 +/- 9.74	313.35 +/- 92.16	278.54 +/- 85.71			110.83 +/- 44.44	101.22 +/- 33.57	Interaction after	12.4 +/- 7.64	12.3 8.87
Correlation	-0.160 (0.541)		-0.045 (0.865)		+0.260 (0.297)				+0.086 (0.836)			-0.012 (0.961)	
.p=	0.246		0.007		0.248				0.808			0.955	

Analysis of correlations and significant differences between test and retest for the test population.

Group 1 Test Retest Analysis	Attention Test		Discrimination Test		Social Cues Test		Novel Object Test-Block Interaction		Novel Environment Test		Bin Interaction	Social Sensitivity Test	
	Test	Retest	Test	Retest	Test	Retest	Test	Retest	Test	Retest		Test	Retest
Correct	28.1 +/- 6.6	24.2 +/-6	35.4 +/- 15.2	42.7 +/- 10.4	2.55 +/- 1.97	1.44 +/- 2.07	Block Sniffing time 1.38 +/- 1.44	Block Sniffing time 2.42 +/-2.28	8.64 +/- 4.7	5.9 +/-5.5	Latency to interact before voice	3.85 +/- 6.42	2.2 +/- 2.26
Correlation	+0.355 (0.349)		+0.504 (0.249)		+0.152 (0.696)		+0.265(0.459)		+0.504 (0.137)			-0.228 (0.420)	
.p=	0.042		0.304		0.204		0.138		0.108			0.878	
Incorrect	8 +/- 5.6	4.9 +/- 3.9	0.43 +/- 0.53	0.14 +/- 0.38	11.55 +/- 2.84	11.67 +/- 5.58	Bin sniffing time 0.15 +/- 0.37	Bin sniffing time 0.46 +/-0.83			Interaction before voice	8.91 +/- 9.01	8.3 +/- 11.13
Correlation	-0.410 (0.273)		0.471 (0.286)		-0.476 (0.196)		-0.172 (0.634)					-0.588 (0.074)	
.p=	0.212		0.157		0.889		0.345					0.878	
Latency to switch or interact	64.7 +/- 46.1	35.6 +/- 29.5	1.13 +/- 2.3	7.78 +/-17							Startle	1.51 +/- 1.56	1.38 +/- 1.91
Correlation	-0.618 (0.076)		+0.883 (**0.002)									-0.049 (0.892)	
.p=	0.173		0.225									0.779	
Opt out	5.8 +/- 5.3	2.7 +/- 3.0			3.36 +/- 1.86	4.11 +/- 3.44					Door directed behaviour	7.25 +/- 12.54	5.1 +/- 8.13
Correlation	-0.604 (0.085)				+0.014 (0.971)							+0.088 (0.809)	
.p=	0.192				0.483							0.779	
Investigation									25.35 +/- 26.61	16.52 +/- 18.7	Latency to interact after	8.36 +/- 9.86	9.34 +/- 10.77
Correlation									-0.017 (0.962)			-0.646 (*0.043)	
.p=									0.507			0.878	
Trials to learn or total time	317.4 +/- 162.9	269.7 +/- 222.6	44.4 +/- 26.6	34.32 +/- 11.2	321.18 +/- 94.6	277.22 +/- 91.7			118.78 +/- 46.16	112.78 +/- 35.85	Interaction after	15.09 +/- 8.13	12.5 +/- 7.98
Correlation	-0.549 (0.126)		-0.412 (0.359)		-0.278 (0.468)				+0.204 (0.573)			-0.294 (0.409)	
.p=	0.515		0.553		0.441				0.959			0.507	

Analysis of correlations and significant differences between test and retest for group 1 (Optimal Group).

Group 2 Test Retest Analysis	Attention Test		Discrimination Test		Social Cues Test		Novel Object Test-Block Interaction		Novel Environment Test		Bin Interaction	Social Sensitivity Test	
	Test	Retest	Test	Retest	Test	Retest	Test	Retest	Test	Retest		Test	Retest
Correct	31.75 +/- 14.05	27.2 +/- 9.11	24.3 +/- 16.89	25.7 +/- 13.6	1.6 +/- 0.97	1.89 +/- 2.57	Block Sniffing time 1.72 +/- 2.06	Block Sniffing time 1.38 +/- 2.53	7.5 +/- 6.98	3.2 +/- 3.55	Latency to interact before voice	6.1 +/- 6.17	6.48 +/- 8.36
Correlation	+0.578 (0.134)		+0.264 (0.462)		+0.227 (0.557)		+0.794 (**0.006)		+0.22 (0.542)			-0.215 (0.551)	
.p=	0.123		0.799		0.831		0.497		0.042			0.610	
Incorrect	7.6 +/- 4.53	6.8 +/- 4.13	1 +/- 1.41	1 +/- 0.82	9.9 +/- 4.86	9.89 +/- 4.34	Bin sniffing time 0.38 +/- 0.77	Bin sniffing time 0.36 +/- 0.41			Interaction before voice	6 +/- 7.80	4.8 +/- 6.44
Correlation	+0.417 (0.203)		+0.385 (0.272)		+0.407 (0.277)		+0.442 (0.225)					+0.349 (0.323)	
.p=	0.672		0.931		0.489		1					0.594	
Latency to switch or interact	44.36 +/- 37.25	31.34 +/- 15.74	19.42 +/- 33.72	8.2 +/- 16.29							Startle	2.54 +/- 4.51	0.58 +/- 1.04
Correlation	-0.035 (0.924)		+0.867 (**0.001)									+0.310 (0.383)	
.p=	0.385		0.207									0.173	
Opt out	5 +/- 4.29	2 +/- 2.16			4.9 +/- 2.56	5.3 +/- 3.74					Door directed behaviour	5.38 +/- 4.24	8.42 +/- 8.36
Correlation	-0.06 (0.869)				-0.376 (0.318)							+0.683 (0.615)	
.p=	0.097				0.528							0.683	
Investigation									14.96 +/- 15.88	5.9 +/- 6.49	Latency to interact after	10.38 +/- 10.4	9.54 +/- 9.6
Correlation									+0.219 (0.543)			+0.123 (0.735)	
.p=									0.066			0.859	
Trials to learn or total time	256.01 +/- 142.2	222.16 +/- 129.64	88.6 +/- 42.37	31.8 +/- 9.08	304.74 +/- 93.67	219.87 +/- 84.86			102.08 +/- 43.12	89.66 +/- 28.22	Interaction after	9.44 +/- 6.14	12.1 +/- 10.11
Correlation	+0.499 (0.208)		+0.258 (0.471)		+0.799 (**0.010)				-0.397 (0.257)			+0.274 (0.444)	
.p=	0.484		0.009		0.214				0.646			0.508	

Analysis of correlations and significant differences between test and retest for group 2 (Lower Scoring Group).